Probability and Markov Chains

9.1 Probability: An Informal Approach

Our approach to probability will be very basic in this section; we will be more formal in the next section.

We begin by considering a *set*; recall that a *set* can be thought of as a *collection of objects.* For example, consider a deck of regular playing cards, consisting of 52 cards. This set will be called the *sample space* or the *universe*. Now suppose we shuffle the deck a number of times and, at random, pick out a card. Assume that the card is the King of Diamonds. The action of selecting this card is called an *event.* And we might ask the following question: "How likely are we to pick out the King of Diamonds?"

Before attempting to answer *this* question, let us consider the following:

- How many times have we shuffled the deck?
- What do we mean by random?
- What do we mean by likely?

The answers to these three "simple" questions touch on a number of advanced mathematical concepts and can even go into philosophical areas. For our purposes, we will, by and large, appeal to our intuition when quantifying certain concepts beyond the scope of this section.

However, we can give a *reasonable* answer to our original question. We note that the *size* of our sample space (the deck of cards) is 52. We also observe that there is only one way to draw a King of Diamonds, since there is *only one* King of Diamonds in the deck; hence, the *size* of the desired event is 1. So we make the following statement, which should seem plausible to the reader: we will say that *the probability of the desired event is, simply, 1/52*.

Using mathematical notation, we can let S be the set that represents the sample space and let E be the set that represents the desired event. Since the number of objects (size) in any set is called the *cardinal number*, we can write $N(S) = 52$ and $N(E) = 1$, to represent the cardinal number of each set. So we now write

$$P(E) = \frac{N(E)}{N(S)} = \frac{1}{52} \tag{1}$$

to denote the *probability of event E*.

What does this mean? It does *not* mean that, should we make exactly 52 drawings of a card—returning it to the deck after each drawing and reshuffling the deck each time—we would draw the King of Diamonds exactly once. (Try it!).

A better interpretation is that *over a very large number of trials*, the *proportion* of times for which a King of Diamonds would be drawn, would get closer and closer to 1 out of 52.

Continuing with this example, the probability of drawing a Spade (event F) is one-fourth, since there are 13 Spades in the deck,

$$P(F) = \frac{N(F)}{N(S)} = \frac{13}{52} = \frac{1}{4}. \tag{2}$$

Another example would be to consider a fair die; let's call it D. Since there are six faces, there are six equally likely outcomes (1, 2, 3, 4, 5, or 6) for every roll of the die, $N(D) = 6$. If the event G is to roll a "3," then

$$P(G) = \frac{N(G)}{N(D)} = \frac{1}{6}. \tag{3}$$

Experiment by rolling a die "many" times. You will find that the proportion of times a "3" occurs is close to one-sixth. In fact, if this is *not* the case, the die is most probably "not fair".

Remark 1 From this example it is clear that the probability of any of the six outcomes is one-sixth. Note, too, that the *sum* of the six probabilities is 1. Also, the probability of rolling a "7" is *zero*, simply because there are no 7s on the any of the six faces of the die.

Because of the above examples, it is most natural to think of probability as a *number.* This number will always be between 0 and 1. We say that an event is *certain* if the probability is 1, and that it is *impossible* if the probability is 0. Most probabilities will be strictly between 0 and 1.

To *compute* the number we call the *probability of an event*, we will adopt the following convention. *We will divide the number of ways the desired event can occur by the total number of possible outcomes. We always assume that each member of the sample space is "just as likely" to occur as any other member.* We call this a *relative frequency* approach.

Example 1 Consider a fair die. Find the probability of rolling a number that is a perfect square.

As before, the size of the sample space, D, is 6. The number of ways a perfect square can occur is two: only "1" or "4" are perfect squares out of the first six positive integers. Therefore, the desired probability is

$$P(K) = \frac{N(K)}{N(D)} = \frac{2}{6} = \frac{1}{3}. \tag{4}$$

■

Example 2 Consider a pair of fair dice. What is the probability of rolling a "7"?

To solve this problem, we first have to find the cardinal number of the sample space, R. To do this, it may be helpful to consider the dice as composed of one red die and one green die, and to think of a "roll" as tossing the red die first, followed by the green die. Then $N(R) = 36$, because there are 36 possible outcomes. To see this, consider Figure 9.1 below. Here, the first column represents the outcome of the red die, the first row represents the outcome of the green die and the body is the *sum* of the two dice—the actual number obtained by the dice roll.

Notice, too, that if we label a "7" roll event Z, then $N(Z) = 6$, because there are six distinct ways of rolling a "7"; again, see Figure 9.1 below.

So our answer is

$$P(Z) = \frac{N(Z)}{N(R)} = \frac{6}{36} = \frac{1}{6}. \tag{5}$$

■

Example 3 Suppose a random number generator generates numbers ranging from 1 through 1000. Find the probability that a given number is divisible by 5.

Elementary Number Theory teaches that for a number to be divisible by 5, the number must end in either a "5" or a "0". By sheer counting, we know that there are 200 numbers between 1 and 1000 that satisfy this condition. Therefore, the required probability is $\frac{200}{1000}$. ■

R\G	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure 9.1

In the next section we will give some rules that pertain to probabilities, and investigate the meaning of probability more fully.

Problems 9.1

1. Find the sample space, its cardinal number and the probability of the desired event for each of the following scenarios:

a) Pick a letter, at random, out of the English alphabet. Desired Event: choosing a vowel.

b) Pick a date, at random, for the Calendar Year 2008. Desired Event: choosing December 7th.

c) Pick a U.S. President, at random, from a list of all the presidents. Desired Event: choosing Abraham Lincoln.

d) Pick a U.S. President, at random, from a list of all the presidents. Desired Event: choosing Grover Cleveland.

e) Pick a card, at random, from a well shuffled deck of regular playing cards. Desired Event: choosing the Ace of Spades.

f) Pick a card, at random, from a well shuffled deck of Pinochle playing cards. Desired Event: choosing the Ace of Spades.

g) Roll a pair of fair dice. Desired Event: getting a roll of "2" (Snake Eyes).

h) Roll a pair of fair dice. Desired Event: getting a roll of "12" (Box Cars).

i) Roll a pair of fair dice. Desired Event: getting a roll of "8".

j) Roll a pair of fair dice. Desired Event: getting a roll of "11".

k) Roll a pair of fair dice. Desired Event: getting a roll of an even number.

l) Roll a pair of fair dice. Desired Event: getting a roll of a number that is a perfect square.

m) Roll a pair of fair dice. Desired Event: getting a roll of a number that is a perfect cube.

n) Roll a pair of fair dice. Desired Event: getting a roll of a number that is a multiple of 3.

o) Roll a pair of fair dice. Desired Event: getting a roll of a number that is divisible by 3.

p) Roll a pair of fair dice. Desired Event: getting a roll of "13".

2. Suppose we were to roll three fair dice: a red one first, followed by a white die, followed by a blue die. Describe the sample space and find its cardinal number.

3. Suppose the probability for event A is known to be 0.4. Find the cardinal number of the sample space if $N(A) = 36$.

4. Suppose the probability for event B is known to be 0.65. Find the cardinal number of B, if the cardinal number of the sample space, S, is $N(S) = 3000$.

9.2 Some Laws of Probability

In this section we will continue our discussion of probability from a more theoretical and formal perspective.

Recall that the probability of event A, given a sample space S, is given by

$$P(A) = \frac{N(A)}{N(S)}, \tag{6}$$

where the numerator and denominator are the respective cardinal numbers of A and S.

We will now give a number of definitions and rules which we will follow regarding the computations of probabilities. This list "formalizes" our approach. We note that the reader can find the mathematical justification in any number of sources devoted to more advanced treatments of this topic. We assume that A, B, $C\ldots$ are any events and that S is the sample space. We also use Φ to denote an impossible event.

- $P(\Phi) = 0$; that is, the probability of an impossible event is zero.
- $P(S) = 1$; that is, the probability of a certain event is one.
- For any two events, A and B, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{7}$$

Remark 1 Here we use both the *union* ($\cup$) and *intersection* ($\cap$) notation from set theory. For this rule, we subtract off the probability of the "common" even in order not to "count it twice". For example, if A is the event of drawing a *King* from a deck of regular playing cards, and B is the event of drawing a Diamond, clearly the *King of Diamonds* is both a King *and* a Diamond. Since there are 4 Kings in the deck, the probability of drawing a King is $\frac{4}{52}$. And since there are 13 Diamonds in the deck, the probability of drawing a Diamond is $\frac{13}{52}$. Since there is only one King of Diamonds in the deck, the probability of drawing this card is clearly $\frac{1}{52}$. We note that

$$\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}, \tag{8}$$

which is the probability of drawing a King *or* a Diamond, because the deck contains *only* 16 Kings *or* Diamonds.

- If A and B are disjoint events, then

$$P(A \cup B) = P(A) + P(B). \tag{9}$$

Remark 2 Two events are *disjoint* if they are mutually exclusive; that is, they cannot happen simultaneously. For example, the events of drawing a *King* from a deck of regular playing cards and, at the same time, drawing a *Queen* are disjoint events. In this case, we merely add the individual probabilities. Note, also, that since A and B are disjoint, we can write $A \cap B = \Phi$; hence, $P(A \cap B) = P(\Phi) = 0$. The reader will also see that equation (9) is merely a special case of equation (7).

- Consider event A; if A^C represents the *complement of* A, then

$$P(A^C) = 1 - P(A). \tag{10}$$

Remark 3 This follows from the fact that either an event occurs or it doesn't. Therefore, $P(A \cup A^C) = 1$; but since these two events are disjoint, $P(A \cup A^C) = P(A) + P(A^C) = 1$. Equation (10) follows directly.

For example, if the probability of rolling a "3" on a fair die is $\frac{1}{6}$, then the probability of *not* rolling a "3" is $1 - \frac{1}{6} = \frac{5}{6}$.

In the next section we will introduce the idea of *independent events*, along with associated concepts. For the rest of this section, we give a number of examples regarding the above rules of probability.

Example 1 Given a pair of fair dice, find the probability of rolling a "3" or a "4".

Since these events are disjoint, we use Equation (9) and refer to Figure 9.1 and obtain the desired probability: $\frac{2}{36} + \frac{3}{36} = \frac{5}{36}$. ■

Example 2 Given a pair of fair dice, find the probability of not rolling a "3" or a "4".

From the previous example, we know that the probability of rolling a "3" or a "4" is $\frac{5}{36}$, therefore, using Equation (10), we find that the probability of the complementary event is: $1 - \frac{5}{36} = \frac{31}{36}$. ■

Remark 4 Note that we could have computed this probability directly by counting the number of ways —31— in which the rolls 2, 5, 6, 7, 8, 9, 10, 11 or 12 can occur. However, using Equation (10) is the preferred method because it is quicker.

Example 3 Pick a card at random out of a well shuffled deck of regular playing cards. What is the probability of drawing a picture card (that is, a King, Queen, or Jack)?

Since there are four suits (Spades, Hearts, Clubs, and Diamonds), and there are three picture cards for each suit, the desired event can occur 12 ways; these can be thought of as 12 disjoint events. Hence, the required probability is $\frac{12}{52}$. ■

Example 4 Pick a card at random out of a well shuffled deck of regular playing cards. Find the probability of drawing a red card or a picture card.

We know there are 12 picture cards as well as 26 red cards (Hearts or Diamonds). But our events are not disjoint, since six of the picture cards are red. Therefore, we apply Equation (7) and compute the desired probability as $\frac{12}{52} + \frac{26}{52} - \frac{6}{52} = \frac{32}{52}$. ■

Example 5 Suppose events A and B are not disjoint. Find $P(A \cap B)$, if it is given that $P(A) = 0.4$, $P(B) = 0.3$, and $P(A \cup B) = 0.6$.

Recall Equation (7): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Therefore, $0.6 = 0.4 + 0.3 - P(A \cap B)$. Therefore, $P(A \cap B) = 0.1$. ■

Example 6 Extend formula (7) for three non-disjoint events. That is, consider $P(A \cup B \cup C)$.

By using parentheses to group two events, we have the following equation below:

$$P(A \cup B \cup C) = P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B)) \cap C). \quad (11)$$

From Set Theory, we know that the last term of (11) can be written as $P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C))$. Hence, applying (7) to (11) yields

$$P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C)). \quad (12)$$

But the last term of (12) is equivalent to $P(A \cap B \cap C)$. After applying (7) to the $P(A \cup B)$ term in (11), we have

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \\ - [P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C))]. \quad (13)$$

This simplifies to:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ - P(B \cap C) + P(A \cap B \cap C). \quad (14)$$

■

Remark 5 Equation (14) can be extended for any finite number of events, and it holds even if some events are pairwise disjoint. For example, if events A and B are disjoint, we merely substitute $P(A \cap B) = 0$ into (14).

Problems 9.2

1. Pick a card at random from a well shuffled deck of regular playing cards. Find the probabilities of:

a) Picking an Ace or a King.

b) Picking an Ace or a picture card.

c) Picking an Ace or a black card.

d) Picking the Four of Diamonds or the Six of Clubs.

e) Picking a red card or a Deuce.

f) Picking a Heart or a Spade.

g) Not choosing a Diamond.

h) Not choosing a Queen.

i) Not choosing an Ace or a Spade.

2. Roll a pair of fair dice. Find the probabilities of:

a) Getting an odd number.

b) Rolling a prime number.

c) Rolling a number divisible by four.

d) Not rolling a "7".

e) Not rolling a "6" or an "8".

f) Not rolling a "1".

3. See Problem 2 of Section 9.1. Roll the three dice. Find the probabilities of:

a) Getting an odd number.

b) Rolling a "3".

c) Rolling an "18".

d) Rolling a "4".

e) Rolling a "17".

f) Rolling a "25".

g) Not rolling a "4".

h) Not rolling a "5".

i) Not rolling a "6".

4. Consider events A and B. Given $P(A) = .7$ and $P(B) = .2$, find the probability of "A or B" if $P(A \cap B) = .15$.

5. Suppose events A and B are equally likely. Find their probabilities if $P(A \cup B) = .46$ and $P(A \cap B) = .34$.

6. Extend Equation (14) for any four events A, B, C, and D.

9.3 Bernoulli Trials and Combinatorics

In the previous section, we considered single events. For example, rolling dice *once* or drawing *one* card out of a deck. In this section we consider multiple events which neither affect nor are affected by preceding or succeeding events.

For this section, we will consider events with only *two* outcomes. For example, flipping a coin, which can result in only "heads" or "tails". The coin cannot land on its edge. We do not insist that the probability of a "head" equals the probability of a "tail", but we will assume that the probabilities remain constant. Each one of the "flips" will be called a *Bernoulli trial*, in honor of Jakob Bernoulli (1654–1705).

Remark 1 A closely related underlying mathematical structure for these trials is known as a *Binomial Distribution*.

Remark 2 The Bernoulli family had a number of great mathematicians and scientists spanning several generations. This family is to mathematics what the Bach family is to music.

As we have indicated, we will assume that the events are *independent*. Hence the probabilities are unaffected at all times. So, if we tossed a coin 10 times in a row, each of the tosses would be called a *Bernoulli trial*, and the probability of getting a head on each toss would remain constant.

We will assume the following rule. If two events, A and B, are independent, then the probability of "*A and B*" or the probability of "*A followed by B*" is given by:

$$P(AB) = P(A \cap B) = P(A)P(B). \tag{15}$$

Notice that we use the intersection ($\cap$) notation. This simple rule is called the *multiplication rule*.

Remark 3 The reader must be careful not to confuse *disjoint* events with *independent* events. The former means that "nothing is in common" or that the events "cannot happen simultaneously". The latter means that the probabilities do not influence one another. Often, but not always, independent events are *sequential*; like flipping a coin 10 times in a row.

It is clear that probabilities depend on *counting*, as in determining the size of the sample space. We assume the following result from an area of mathematics known as *Combinatorics*:

- The number of ways we can choose k objects from a given collection of n objects is given by:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{16}$$

Remark 4 This is equivalent to determining the number of *subsets of size k* given a *set of size n*, where $k \leq n$.

Remark 5 We saw "factorials" in Chapter Seven. Recall that *3!*, for example, is read "three factorial" and it is evaluated $3 \cdot 2 \cdot 1 = 6$. Hence, "n − factorial" is given by $n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1$. By convention, we define $0! = 1$. Finally, we only consider cases where n is a non-negative integer.

Remark 6 For these "number of ways", we are not concerned about the *order* of selection. We are merely interested in the number of *combinations* (as opposed to the number of *permutations*).

We will provide the reader with a number of examples which illustrate the salient points of this section.

Example 1 Evaluate $\binom{5}{2}$. Using (16) we see that $\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5!}{2!3!}$.
Since $5! = 120$, $2! = 2$ and $3! = 6$, $\binom{5}{2} = \frac{120}{12} = 10$. ■

Example 2 Given a committee of five people, in how many ways can a sub-committee of size two be formed? The number is precisely what we computed in the previous example: 10. The reader can verify this as follows. Suppose the people are designated: A, B, C, D, and E. Then, the 10 sub-committees of size two are given by: AB, AC, AD, AE, BC, BD, BE, CD, CE, and DE. ■

Example 3 Given a committee of five people, in how many ways can a sub-committee of size three be formed? We can use formula (16) to compute the answer, however the answer must be 10. This is because a sub-committee of 3 is the *complement* of a sub-committee or two. That is, consider the three people *not* on a particular sub-committee of two, as constituting a sub-committee of three. For example, if A and B are on one sub-committee of size two, put C, D, and E on a sub-committee of size three. Clearly there are 10 such pairings. ■

Example 4 Suppose we flip a fair coin twice. What is the probability of getting *exactly* one "head"?

Let H represent getting a "head" and T represent getting a "tail". Since the coin is fair, $P(H) = P(T) = \frac{1}{2}$. The only way we can obtain exactly *one* head in *two* tosses is if the order of outcomes is either HT or TH. Note that the events are *disjoint* or *mutually exclusive*; that is, we cannot get these two outcomes at the same time. Hence, Equation (9) will come into play. And because of the *independence* of the coin flips (each toss is a *Bernoulli trial*), we will use Equation (15) to determine the probability of obtaining HT and TH.

Therefore, the probability of getting exactly one H is equal to

$$\begin{aligned} P(HT \cup TH) &= P(HT) + P(TH) = P(H)P(T) + P(T)P(H) \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}. \end{aligned} \tag{17}$$

■

Remark 7 Note that (17) could have been obtained by finding the probability of HT—in that order—and then multiplying it by the *number of times* (*combinations*) we could get exactly one H in two tosses.

Example 5 Now, suppose we flip a fair coin 10 times. What is the probability of getting exactly one "head"?

Suppose we get the H on the first toss. Then the probability of getting $HTTTTTTTTT$—in that order—is equal to $\left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right)^9 = \frac{1}{1024}$, because the tosses are all independent. Note that if the H occurs "in the second slot", the probability is also $\frac{1}{1024}$. In fact, we get the same number for all 10 possible "slots". Hence the final answer to our question is $\frac{10}{1024}$. ■

Remark 8 Note that the previous example could have be answered by the following computation

$$\binom{10}{1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^9 = \frac{10}{1024}. \tag{18}$$

Here, the first factor gives the number of ways we can get exactly one H in 10 tosses; this is where mutual exclusivity comes in. The second factor is the probability of getting one H, and the third factor is the probability of getting nine Ts; the independence factor is employed here.

Example 6 Suppose we flip a fair coin 10 times. Find the probability of getting *exactly* five Hs.

Since there are $\binom{10}{5} = 252$ ways of getting five Hs in 10 tosses, the desired probability is given by

$$\binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{252}{1024}. \tag{19}$$

■

Example 7 Suppose we flip an *unfair* coin 10 times. If $P(H) = .3$ and the $P(T) = .7$, what is the probability of getting *exactly* five Hs?

As we learned from the previous problem, there are 252 ways of getting exactly five Hs in 10 tosses. Hence, our desired probability is given by

$$\binom{10}{5}(.3)^5(.7)^5 \approx 0.103. \quad ■$$

Remark 9 A calculator is useful for numerical computations. We will address the issues of calculations and technology in both Section 9.5 and in the Appendix. Note, too, that individual probabilities, $P(H)$ and $P(T)$ must add to 1, and that the exponents in the formula must add to the total number of tosses; in this case, 10.

Example 8 Consider the previous example. Find the probability of getting *at least five Hs*.

Note that *at least five Hs* means *exactly five Hs* plus *exactly six Hs* plus … etc. Note, also, that *exactly five Hs* and *exactly six Hs* are disjoint, so we will use Equation (9). Therefore the desired probability is given by:

$$\binom{10}{5}(.3)^5(.7)^5 + \binom{10}{6}(.3)^6(.7)^4 + \binom{10}{7}(.3)^7(.7)^3 + \binom{10}{8}(.3)^8(.7)^2$$
$$+ \binom{10}{9}(.3)^9(.7)^1 + \binom{10}{10}(.3)^{10}(.7)^{10} \approx 0.150. \tag{20}$$

■

Example 9 Consider the previous examples. Find the probability of getting *at least one H*.

While we could follow the approach in Example 9, there is a simpler way to answer this question. If we realize that the *complement* of the desired event is *getting no Hs* in 10 tosses, we can apply Equation (10). That is, the probability of getting at *least one H* is equal to

$$1 - \binom{10}{0}(.3)^0(.7)^{10} \approx 0.972. \tag{21}$$

■

We summarize the results for the probability in this section as follows:

- Given n successive Bernoulli trials, the probability of getting exactly k successes, where $k \leq n$, is equal to

$$\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \tag{22}$$

where the probability of a success is p, and the probability of a failure is $(1-p)$.

Problems 9.3

1. Evaluate the following:

a) $\binom{6}{2}$; b) $\binom{7}{1}$; c) $\binom{8}{5}$; d) $\binom{20}{18}$; e) $\binom{20}{2}$;

f) $\binom{1000}{1000}$; g) $\binom{1000}{0}$; h) $\binom{100}{99}$; i) $\binom{1000}{999}$; j) $\binom{0}{0}$.

2. How many different nine-player line-ups can the New York Yankees produce if there are 25 players on the roster and every player can play every position?

3. Suppose 15 women comprised a club, and a committee of 6 members was needed. How many different committees would be possible?

4. Toss a fair die eight times. Find the probability of:

a) Rolling exactly one "5".

b) Rolling exactly three "5s".

c) Rolling at least three "5s".

d) Rolling at most three "5s".

e) Rolling at least one "5".

5. Suppose event A has a probability of occurring equal to .65. Without evaluating the expressions, find the following probabilities given 500 independent Bernoulli trials.

a) Event A occurs 123 times.

b) Event A occurs 485 times.

c) Event A occurs at least 497 times.

d) Event A occurs at most 497 times.

e) Event A occurs any non-zero multiple of 100 times.

6. An urn contains 10 golf balls, three of which are white, with the remaining seven painted orange. A blindfolded golfer picks a golf ball from the urn, and then replaces it. The process is repeated nine times, making a total of 10 trials. What is the probability of the golfer picking a white golf ball exactly three times?

7. An urn contains 10 golf balls colored as follows: three are white; two are green; one is red; four are orange. A blindfolded golfer picks a golf ball from the urn, and then replaces it. The process is repeated nine times, making a total of 10 trials. What is the probability of the golfer picking a white golf ball exactly three times?

9.4 Modeling with Markov Chains: An Introduction

In Chapter 1 and Chapter 6, we mentioned the concept of Markov chains. We return to this idea, further formalizing it from the perspective of probability. Consider the following example, which we will call the Moving Situation.

Example 1 Suppose we have two families. Family (1) lives in state A and Family (2) lives in state B. Let us further assume that the matrix

$$P = \begin{bmatrix} .7 & .3 \\ .9 & .1 \end{bmatrix} \tag{23}$$

represents the following probabilities. The element in the first row and first column represents the probability of Family (1) originally residing in state A remaining in state A, while the element in the first row and second column represents the probability of starting in state A and then moving to state B. Note that these two probabilities add to one.

Similarly, let the element in the second row and first column, represent the probability of Family (2) starting in state B and moving to state A, while the element in the second row and second column, represents the probability of starting in state B and remaining in state B. Here, too, these two probabilities add to one.

Note that we can consider the process as "time conditioned" in the sense that there is a *present* and a *future* (for example, one year from the present).

Such a matrix is called a *transition matrix* and the elements are called *transitional probabilities.*

Let us consider the matrix in (23) and let us compute P^2. We find that

$$P^2 = \begin{bmatrix} .76 & .24 \\ .72 & .28 \end{bmatrix}. \tag{24}$$

What does P^2 represent? To answer this question, let us ask another question: From the perspective of Family (1), what is the probability of being in state A after *two* years?

There are two ways Family (1) can be in state A after two years:

- Scenario 1: Either the family stayed two years in a row.
- Scenario 2: The family moved to state B after one year and then moved back to state A after the second year.

The probability of the first scenario is $.7(.7) = .49$, because these events can be considered *independent*. The probability of the second is $.3(.7) = .27$.

Because these events are *disjoint*, we add the probabilities to get .76.

Note that this is the element in the first row and first column of P^2.

By similar analyses we find that P^2 is indeed the transitional matrix of our Moving Situation after two time periods.

Matrix P is the transition matrix for a *Markov chain*. The sum of the probabilities of each row must add to one, and by the very nature of the process, the matrix must be square. We assume that at any time each object is in one and only one state (although different objects can be in the same state). We also assume that the probabilities remain constant over the given time period. ■

Remark 1 The notation $p^{(n)}{}_{ij}$ is used to signify the transitional probability of moving from state i to state j over n time periods.

Example 2 Suppose Moe, Curly, and Larry live in the same neighborhood. Let the transition matrix

$$S = \begin{bmatrix} .7 & .1 & .2 \\ .5 & .3 & .2 \\ .8 & .1 & .1 \end{bmatrix} \tag{25}$$

represent the respective probabilities of Moe, Curly, and Larry staying at home on Monday and either visiting one of their two neighbors or staying home on Tuesday. We ask the following questions regarding Thursday:

a) What is the probability of Moe going to visit Larry at his home, $p^{(3)}{}_{13}$?

b) What is the probability of Curly being at his own home, $p^{(3)}{}_{22}$?

To answer both of these questions, we must compute P^3 because three time periods would have elapsed. We find that

$$P^3 = \begin{bmatrix} .694 & .124 & .182 \\ .124 & .132 & .182 \\ .695 & .124 & .181 \end{bmatrix}. \tag{26}$$

So, our answers are as follows: a) the probability is .182, the entry in the first row and third column; b) the probability is .132, the entry in the second row and second column. ■

Example 3 Consider the transitional matrix K which represents respective probabilities of Republicans, Democrats, and Independents either remaining within their political parties or changing their political parties over a two-year period:

$$K = \begin{bmatrix} .7 & .1 & .2 \\ .15 & .75 & .1 \\ .3 & .2 & .5 \end{bmatrix}. \tag{27}$$

What is the probability of a Republican becoming an Independent after four years? And what is the probability of a Democrat becoming a Republican after four years?

Both of these questions require *two* time periods; hence we need $p^{(2)}{}_{13}$ and $p^{(2)}{}_{21}$ which can be obtained from K^2 below:

$$K^2 = \begin{bmatrix} .565 & .185 & .25 \\ .2475 & .5975 & .155 \\ .39 & .28 & .33 \end{bmatrix}. \tag{28}$$

Hence, $p^{(2)}{}_{13} = .25$ and $p^{(2)}{}_{21} = .2475$.

We close this discussion with the observation that sometimes transitional matrices have special properties. For example, consider the matrix

$$A = \begin{bmatrix} .5 & .5 \\ .3 & .7 \end{bmatrix}. \tag{29}$$

We find that if we raise this matrix to the 9th and 10th powers, our result is the same. That is,

$$A^9 = A^{10} = \begin{bmatrix} .375 & .625 \\ .375 & .625 \end{bmatrix}. \tag{30}$$

and the *same result occurs for any higher power of A*. This *absorbing* quality implies that "sooner or later" the transitional probabilities will stabilize.

Markov processes are used in many areas including decision theory, economics and political science. ■

Problems 9.4

1. Why are the following matrices not transitional matrices?

a) $\begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$; b) $\begin{bmatrix} .6 & .5 \\ .4 & .5 \end{bmatrix}$; c) $\begin{bmatrix} .1 & .2 & .7 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$; d) $\begin{bmatrix} .1 & .5 & .4 \\ .2 & .6 & .2 \end{bmatrix}$

2. Consider the following transitional matrices. Construct scenarios for which these matrices might represent the transitional probabilities:

a) $\begin{bmatrix} .5 & .5 \\ .7 & .3 \end{bmatrix}$; b) $\begin{bmatrix} .95 & .05 \\ .02 & .98 \end{bmatrix}$; c) $\begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}$;

d) $\begin{bmatrix} 1 & 0 \\ 0 & .1 \end{bmatrix}$; e) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$; f) $\begin{bmatrix} .1 & .2 & .7 \\ .5 & .25 & .25 \\ .3 & .3 & .4 \end{bmatrix}$

3. Consider the c) and d) matrices in the previous problem; show that these matrices are "absorbing" matrices.

4. Consider the e) matrix in Problem (2). Raise the matrix to the powers of 2, 3, 4, 5, 6, 7, and 8. What do you notice? Can you construct a scenario for which this transitional matrix could be a model?

5. Consider the following transitional matrix: $\begin{bmatrix} .6 & .4 \\ .1 & .9 \end{bmatrix}$. Find $p^{(2)}{}_{11}$, $p^{(2)}{}_{21}$, $p^{(3)}{}_{12}$, and $p^{(3)}{}_{22}$.

6. Consider a game called Red-Blue. The rules state that after one "turn" Red can become Blue or remain Red. The same is true with Blue. Suppose you make a bet that after five turns, Red will be in the Red category. You are told that the following probabilities are valid:

- Given Red, the probability of remaining Red after one turn is .7
- Given Red, the probability of going to Blue is .3
- Given Blue, the probability of remaining Blue is .6
- Given Blue, the probability of going to Red is .4

a) Give the transition matrix.

b) What is the probability of you winning your bet?

c) Does the probability increase, decrease or stay the same if you bet six turns instead of five?

9.5 Final Comments on Chapter 9

Probability is a fascinating area. For numbers that necessarily range between 0 and 1, inclusively, a lot can happen.

When using probability, we must understand exactly what is being asked and give precise answers, without misrepresenting our conclusions. Concepts such as randomness and independence must be present before certain laws can be applied. While the mathematical underpinnings are rock solid, probabilities generally deal with "trends" and "likelihoods".

Regarding Bernoulli trials, if the number of experiments is *large*, the calculations can be overwhelming. In these cases, the use of computers and other technological aids is essential. From a theoretical perspective, there is a very good approximation that can be employed, known as the *Normal Approximation to the Binomial Distribution.* This technique is explored in basic courses on probability and statistics.

One final point: With the exception of the section on Markov chains, all the probabilities in Chapter 9 were *theoretically* assigned. That is, we made assumptions, applied definitions and then made our computations. For example, *if* a die was *fair*, *then* we assigned a probability of $\frac{1}{6}$ to the event of rolling a "3", based on our definition, which dealt with relative frequency.

However, there are many times when probabilities are obtained by *observation* and *empirical evidence*. For example, the greatest baseball player of all time, Babe Ruth, had a lifetime batting average of .342. Since batting average is defined as successful hits divided by total at-bats, we can interpret this as Ruth getting 342 hits for every 1000 at-bats *over a long period of time*.

There are many other occurrences of *empirical probabilities* in research areas such as medicine, psychology, economics, and sociology, to name but a few.