Chapter 6

Probability

6.1   Introduction to Probability

Why do we study probability?

You have studied hashing as a way to store data (or keys to find data) in a way that makes it possible to access that data quickly. Recall that we have a table in which we want to store keys, and we compute a function h of our key to tell us which location (also known as a "slot" or a "bucket") in the table to use for the key. Such a function is chosen with the hope that it will tell us to put different keys in different places, but with the realization that it might not. If the function tells us to put two keys in the same place, we might put them into a linked list that starts at the appropriate place in the table, or we might have some strategy for putting them into some other place in the table itself. If we have a table with a hundred places and fifty keys to put in those places, there is no reason in advance why all fifty of those keys couldn't be assigned (hashed) to the same place in the table. However someone who is experienced with using hash functions and looking at the results will tell you you'd never see this in a million years. On the other hand that same person would also tell you that you'd never see all the keys hash into different locations in a million years either. In fact, it is far less likely that all fifty keys would hash into one place than that all fifty keys would hash into different places, but both events are quite unlikely. Being able to understand just how likely or unlikely such events are is our reason for taking up the study of probability.

In order to assign probabilities to events, we need to have a clear picture of what these events are. Thus we present a model of the kinds of situations in which it is reasonable to assign probabilities, and then recast our questions about probabilities into questions about this model. We use the phrase *sample space* to refer to the set of possible outcomes of a process. For now, we will deal with processes that have finite sample spaces. The process might be a game of cards, a sequence of hashes into a hash table, a sequence of tests on a number to see if it fails to be a prime, a roll of a die, a series of coin flips, a laboratory experiment, a survey, or any of many other possibilities. A set of elements in a sample space is called an *event*. For example, if a professor starts each class with a 3 question true-false quiz the sample space of all possible patterns of correct answers is

$$\{TTT, TTF, TFT, FTT, TFF, FTF, FFT, FFF\}.$$

The event of the first two answers being true is $\{TTT, TTF\}$. In order to compute probabilities we assign a *probability weight* $p(x)$ to each element of the sample space so that the weight represents what we believe to be the relative likelihood of that outcome. There are two rules we must follow in assigning weights. First the weights must be nonnegative numbers, and second the sum of the weights of all the elements in a sample space must be one. We define the *probability* $P(E)$ of the event E to be the sum of the weights of the elements of E. Algebraically we can write

$$P(E) = \sum_{x:x\in E} p(x).$$

We read this as $p(E)$ equals he sum over all x such that x is in E of $p(x)$.

Notice that a probability function P on a sample space S satisfies the rules[1]

1. $P(A) \geq 0$ for any $A \subseteq S$.

2. $P(S) = 1$.

3. $P(A \cup B) = P(A) + P(B)$ for any two disjoint events A and B.

The first two rules reflect our rules for assigning weights above. We say that two events are disjoint if $A \cap B = \emptyset$. The third rule follows directly from the definition of disjoint and our definition of the probability of an event. A function P satisfying these rules is called a *probability distribution* or a *probability measure*.

In the case of the professor's three question quiz, it is natural to expect each sequence of trues and falses to be equally likely. (A professor who showed any pattern of preferences would end up rewarding a student who observed this pattern and used it in educated guessing.) Thus it is natural to assign equal weight $1/8$ to each of the eight elements of our quiz sample space. Then the *probability* of an event E, which we denote by $P(E)$, is the sum of the weights of its elements. Thus the probability of the event "the first answer is T" is $\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$. The event "There is at exactly one True" is $\{TFF, FTF, FFT\}$, so P(there is exactly one True) is $3/8$.

Exercise 6.1-1 Try flipping a coin five times. Did you get at least one head? Repeat five coin flips a few more times! What is the probability of getting at least one head in five flips of a coin? What is the probability of no heads?

Exercise 6.1-2 Find a good sample space for rolling two dice. What weights are appropriate for the members of your sample space? What is the probability of getting a 6 or 7 total on the two dice? Assume the dice are of different colors. What is the probability of getting less than 3 on the red one and more than 3 on the green one?

Exercise 6.1-3 Suppose you hash a list of n keys into a hash table with 20 locations. What is an appropriate sample space, and what is an appropriate weight function? (Assume the keys and the hash function are not in any special relationship to the number 20.) If n is three, what is the probability that all three keys hash to different locations? If you hash ten keys into the table, what is the probability that at least

---

[1]These rules are often called "the axioms of probability." For a finite sample space, we could show that if we started with these axioms, our definition of probability in terms of the weights of individual elements of S is the only definition possible. That is, for any other definition, the probabilities we would compute would still be the same if we take $w(x) = P(\{x\})$.

two keys have hashed to the same location? We say two keys *collide* if they hash to the same location. How big does n have to be to insure that the probability is at least one half that has been at least one collision?

In Exercise 6.1-1 a good sample space is the set of all 5-tuples of Hs and Ts. There are 32 elements in the sample space, and no element has any reason to be more likely than any other, so a natural weight to use is $\frac{1}{32}$ for each element of the sample space. Then the event of at least one head is the set of all elements but $TTTTT$. Since there are 31 elements in this set, its probability is $\frac{31}{32}$. This suggests that you should have observed at least one head pretty often!

The probability of no heads is the weight of the set $\{TTTTT\}$, which is $\frac{1}{32}$. Notice that the probabilities of the event of "no heads" and the opposite event of "at least one head" add to one. This observation suggests a theorem. The *complement* of an event E in a sample space S, denoted by $S - E$, is the set of all outcomes in S but not E. The theorem tells us how to compute the probability of the complement of an event from the probability of the event.

Theorem 6.1 *If two events E and F are complementary, that is they have nothing in common ($E \cap F = \emptyset$) and their union is the whole sample space ($E \cup F = S$), then*

$$P(E) = 1 - P(F).$$

Proof: The sum of all the probabilities of all the elements of the sample space is one, and since we can break this sum into the sum of the probabilities of the elements of E plus the sum of the probabilities of the elements of F, we have

$$P(E) + P(F) = 1,$$

which gives us $P(E) = 1 - P(F)$.∎

For Exercise 6.1-2 a good sample space would be pairs of numbers (a, b) where $(1 \leq a, b \leq 6)$. By the product principle[2], the size of this sample space is $6 \cdot 6 = 36$. Thus a natural weight for each ordered pair is $\frac{1}{36}$. How do we compute the probability of getting a sum of six or seven? There are 5 ways to roll a six and 6 ways to roll a seven, so our event has eleven elements each of weight 1/36. Thus the probability of our event is is 11/36. For the question about the red and green dice, there are two ways for the red one to turn up less than 3, and three ways for the green one to turn up more than 3. Thus, the event of getting less than 3 on the red one and greater than 3 on the green one is a set of size $2 \cdot 3 = 6$ by the product principle. Since each element of the event has weight 1/36, the event has probability 6/36 or 1/6.

In Exercise 6.1-3 an appropriate sample space is the set of n-tuples of numbers between 1 and 20. The first entry in an n-tuple is the position our first key hashes to, the second entry is the position our second key hashes to, and so on. Thus each n tuple represents a possible hash function, and each hash function, applied to our keys, would give us one n-tuple. The size of the sample space is 20^n (why?), so an appropriate weight for an n-tuple is $1/20^n$. To compute the probability of a collision, we will first compute the probability that all keys hash to different locations and then apply Theorem 6.1 which tells us to subtract this probability from 1 to get the probability of collision.

---

[2]from Section 1.1

n	Prob of empty slot	Prob of no collisions
1	1	1
2	0.95	0.95
3	0.9	0.855
4	0.85	0.72675
5	0.8	0.5814
6	0.75	0.43605
7	0.7	0.305235
8	0.65	0.19840275
9	0.6	0.11904165
10	0.55	0.065472908
11	0.5	0.032736454
12	0.45	0.014731404
13	0.4	0.005892562
14	0.35	0.002062397
15	0.3	0.000618719
16	0.25	0.00015468
17	0.2	3.09359E-05
18	0.15	4.64039E-06
19	0.1	4.64039E-07
20	0.05	2.3202E-08

Table 6.1: The probabilities that all elements of a set hash to different entries of a hash table of size 20.

To compute the probability that all keys hash to different locations we consider the event that all keys hash to different locations. This is the set of n tuples in which all the entries are different. (In the terminology of functions, these n-tuples correspond to one-to-one hash functions). There are 20 choices for the first entry of an n-tuple in our event. Since the second entry has to be different, there are 19 choices for the second entry of this n-tuple. Similarly there are 18 choices for the third entry (it has to be different from the first two), 17 for the fourth, and in general $20 - i + 1$ possibilities for the ith entry of the n-tuple. Thus we have

$$20 \cdot 19 \cdot 18 \cdot \cdots \cdot (20 - n + 1) = 20^{\underline{n}}$$

elements of our event.[3] Since each element of this event has weight $1/20^n$, the probability that all the keys hash to different locations is

$$\frac{20 \cdot 19 \cdot 18 \cdot \cdots \cdot (20 - n + 1)}{20^n} = \frac{20^{\underline{n}}}{20^n}.$$

In particular if n is 3 the probability is $(20 \cdot 19 \cdot 18)/20^3 = .855$.

We show the values of this function for n between 0 and 20 in Table 6.1. Note how quickly the probability of getting a collision grows. As you can see with $n = 10$, the probability that there have been no collisions is about .065, so the probability of at least one collision is .935.

If $n = 5$ this number is about .58, and if $n = 6$ this number is about .43. By Theorem 6.1 the probability of a collision is one minus the probability that all the keys hash to different locations.

---

[3]using the notation for falling factorial powers that we introduced in Section 1.2.

Thus if we hash six items into our table, the probability of a collision is more than $1/2$. Our first intuition might well have been that we would need to hash ten items into our table to have probability $1/2$ of a collision. This example shows the importance of supplementing intuition with careful computation!

The technique of computing the probability of an event of interest by first computing the probability of its complementary event and then subtracting from 1 is very useful. You will see many opportunities to use it, perhaps because about half the time it is easier to compute directly the probability that an event doesn't occur than the probability that it does. We stated Theorem 6.1 as a theorem to emphasize the importance of this technique.

The Uniform Probability Distribution

In all three of our exercises it was appropriate to assign the same weight to all members of our sample space. We say P is the *uniform probability measure* or *uniform probability distribution* when we assign the same probability to all members of our sample space. The computations in the exercises suggest another useful theorem.

Theorem 6.2 *Suppose P is the uniform probability measure defined on a sample space S. Then for any event E,*

$$P(E) = |E|/|S|,$$

the size of E divided by the size of S.

Proof: Let $S = \{x_1, x_2, \ldots, x_{|S|}\}$. Since P is the uniform probability measure, there must be some value p such that for each $x_i \in S, P(x_i) = p$. Combining this fact with the second and third probability rules, we obtain

$$
\begin{aligned}
1 &= P(S) \\
&= P(x_1 \cup x_2 \cup \cdots \cup x_{|S|}) \\
&= P(x_1) + P(x_2) + \ldots + P(x_{|S|}) \\
&= p|S| \ .
\end{aligned}
$$

Equivalently

$$p = \frac{1}{|S|} \ . \tag{6.1}$$

E is a subset of S with $|E|$ elements and therefore

$$P(E) = \sum_{x_i \in E} p(x_i) = |E|p \ . \tag{6.2}$$

Combining equations 6.1 and 6.2 gives that $P(E) = |E|p = |E|(1/|S|) = |E|/|S| \ . \blacksquare$

Exercise 6.1-4 What is the probability of an odd number of heads in three tosses of a coin? Use Theorem 6.2.

Using a sample space similar to that of first example (with "T" and "F" replaced by "H" and "T"), we see there are three sequences with one H and there is one sequence with three H's. Thus we have four sequences in the event of "an odd number of heads come up." There are eight sequences in the sample space, so the probability is $\frac{4}{8} = \frac{1}{2}$.

It is comforting that we got one half because of a symmetry inherent in this problem. In flipping coins, heads and tails are equally likely. Further if we are flipping 3 coins, an odd number of heads implies an even number of tails. Therefore, the probability of an odd number of heads, even number of heads, odd number of tails and even number of tails must all be the same. Applying Theorem 6.1 we see that the probability must be 1/2.

A word of caution is appropriate here. Theorem 6.2 applies only to probabilities that come from the equiprobable weighting function. The next example shows that it does not apply in general.

Exercise 6.1-5 A sample space consists of the numbers 0, 1, 2 and 3. We assign weight $\frac{1}{8}$ to 0, $\frac{3}{8}$ to 1, $\frac{3}{8}$ to 2, and $\frac{1}{8}$ to 3. What is the probability that an element of the sample space is positive? Show that this is not the result we would obtain by using the formula of Theorem 6.2.

The event "x is positive" is the set $E = \{1, 2, 3\}$. The probability of E is

$$P(E) = P(1) + P(2) + P(3) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8} \ .$$

However, $\frac{|E|}{|S|} = \frac{3}{4}$.

The previous exercise may seem to be "cooked up" in an unusual way just to prove a point. In fact that sample space and that probability measure could easily arise in studying something as simple as coin flipping.

Exercise 6.1-6 Use the set $\{0, 1, 2, 3\}$ as a sample space for the process of flipping a coin three times and counting the number of heads. Determine the appropriate probability weights $P(0)$, $P(1)$, $P(2)$, and $P(3)$.

There is one way to get the outcome 0, namely tails on each flip. There are, however, three ways to get 1 head and three ways to get two heads. Thus $P(1)$ and $P(2)$ should each be three times $P(0)$. There is one way to get the outcome 3—heads on each flip. Thus $P(3)$ should equal $P(0)$. In equations this gives $P(1) = 3P(0)$, $P(2) = 3P(0)$, and $P(3) = P(0)$. We also have the equation saying all the weights add to one, $P(0) + P(1) + P(2) + P(3) = 1$. There is one and only one solution to these equations, namely $P(0) = \frac{1}{8}$, $P(1) = \frac{3}{8}$, $P(2) = \frac{3}{8}$, and $P(3) = \frac{1}{8}$. Do you notice a relationship between $P(x)$ and the binomial coefficient $\binom{3}{x}$ here? Can you predict the probabilities of 0, 1, 2, 3, and 4 heads in four flips of a coin?

Together, the last two exercises demonstrate that we must be careful not to apply Theorem 6.2 unless we are using the uniform probability measure.

Important Concepts, Formulas, and Theorems

1. *Sample Space.* We use the phrase *sample space* to refer to the set of possible outcomes of a process.

2. *Event.* A set of elements in a sample space is called an *event.*

3. *Probability.* In order to compute probabilities we assign a *weight* to each element of the sample space so that the weight represents what we believe to be the relative likelihood of that outcome. There are two rules we must follow in assigning weights. First the weights must be nonnegative numbers, and second the sum of the weights of all the elements in a sample space must be one. We define the *probability $P(E)$* of the event E to be the sum of the weights of the elements of E.

4. *The axioms of Probability.* Three rules that a probability measure on a finite sample space must satisfy could actually be used to define what we mean by probability.

   (a) $P(A) \geq 0$ for any $A \subseteq S$.
   (b) $P(S) = 1$.
   (c) $P(A \cup B) = P(A) + P(B)$ for any two disjoint events A and B.

5. *Probability Distribution.* A function which assigns a probability to each member of a sample space is called a (discrete) *probability distribution.*

6. *Complement.* The *complement* of an event E in a sample space S, denoted by $S - E$, is the set of all outcomes in S but not E.

7. The Probabilities of Complementary Events. If two events E and F are complementary, that is they have nothing in common ($E \cap F = \emptyset$), and their union is the whole sample space ($E \cup F = S$), then
$$P(E) = 1 - P(F).$$

8. *Collision, Collide (in Hashing).* We say two keys *collide* if they hash to the same location.

9. *Uniform Probability Distribution.* We say P is the *uniform probability measure* or *uniform probability distribution* when we assign the same probability to all members of our sample space.

10. *Computing Probabilities with the Uniform Distribution.* Suppose P is the uniform probability measure defined on a sample space S. Then for any event E,

$$P(E) = |E|/|S|,$$

the size of E divided by the size of S. This *does not* apply to general probability distributions.

Problems

1. What is the probability of exactly three heads when you flip a coin five times? What is the probability of three or more heads when you flip a coin five times?

2. When we roll two dice, what is the probability of getting a sum of 4 or less on the tops?

3. If we hash 3 keys into a hash table with ten slots, what is the probability that all three keys hash to different slots? How big does n have to be so that if we hash n keys to a hash table with 10 slots, the probability is at least a half that some slot has at least two keys hash to it? How many keys do we need to have probability at least two thirds that some slot has at least two keys hash to it?

4. What is the probability of an odd sum when we roll three dice?

5. Suppose we use the numbers 2 through 12 as our sample space for rolling two dice and adding the numbers on top. What would we get for the probability of a sum of 2, 3, or 4, if we used the equiprobable measure on this sample space. Would this make sense?

6. Two pennies, a nickel and a dime are placed in a cup and a first coin and a second coin are drawn.

   (a) Assuming we are sampling without replacement (that is, we don't replace the first coin before taking the second) write down the sample space of all ordered pairs of letters P, N, and D that represent the outcomes. What would you say are the appropriate weights for the elements of the sample space?

   (b) What is the probability of getting eleven cents?

7. Why is the probability of five heads in ten flips of a coin equal to $\frac{63}{256}$?

8. Using 5-element sets as a sample space, determine the probability that a "hand" of 5 cards chosen from an ordinary deck of 52 cards will consist of cards of the same suit.

9. Using 5 element permutations as a sample space, determine the probability that a "hand" of 5 cards chosen from an ordinary deck of 52 cards will have all the cards from the same suit

10. How many five-card hands chosen from a standard deck of playing cards consist of five cards in a row (such as the nine of diamonds, the ten of clubs, jack of clubs, queen of hearts, and king of spades)? Such a hand is called a straight. What is the probability that a five-card hand is a straight? Explore whether you get the same answer by using five element sets as your model of hands or five element permutations as your model of hands.

11. A student taking a ten-question, true-false diagnostic test knows none of the answers and must guess at each one. Compute the probability that the student gets a score of 80 or higher. What is the probability that the grade is 70 or lower?

12. A die is made of a cube with a square painted on one side, a circle on two sides, and a triangle on three sides. If the die is rolled twice, what is the probability that the two shapes we see on top are the same?

13. Are the following two events equally likely? Event 1 consists of drawing an ace and a king when you draw two cards from among the thirteen spades in a deck of cards and event 2 consists of drawing an ace and a king when you draw two cards from the whole deck.

14. There is a retired professor who used to love to go into a probability class of thirty or more students and announce "I will give even money odds that there are two people in this classroom with the same birthday." With thirty students in the room, what is the probability that all have different birthdays? What is the minimum number of students

that must be in the room so that the professor has at least probability one half of winning the bet? What is the probability that he wins his bet if there are 50 students in the room. Does this probability make sense to you? (There is no wrong answer to that question!) Explain why or why not.