Probability
(Devore Chapter Two)

1016-351-70: Probability[*]

Spring 2010

Contents

---

[*]Copyright 2010, John T. Whelan, and all that

Tuesday 16 March 2010

1 Axiomatic Probability

Many of the rules of probability appear to be self-evident, but an important part of mathematics is illustrating that our intuition agrees with the logical outcome of our models. To that end, Devore develops with some care a mathematical theory of probability. We'll mostly summarize the key definitions and results here.

1.1 Outcomes and Events

Devore defines probability in terms of an **experiment** which can have one of a set of possible **outcomes**.

- The **sample space** of an experiment, written $\mathcal{S}$, is the set of all possible outcomes.
- An **event** is a subset of $\mathcal{S}$, a set of possible outcomes to the experiment. Special cases are:
  - The **null event** $\varnothing$ is an event consisting of no outcomes (the empty set)
  - A **simple event** consists of exactly one outcome
  - A **compound event** consists of more than one outcome

The sample space $\mathcal{S}$ itself an event, of course.

One example of an experiment is flipping a coin three times. The outcomes in that case are HHH, HHT, HTH, HTT, THH, THT, TTH, and TTT. Possible outcomes include:

- Exactly two heads: $\{HHT, HTH, THH\}$
- The first flip is heads: $\{HHH, HHT, HTH, HTT\}$
- The second and third flips are the same: $\{HHH, HTT, THH, TTT\}$

Another example is a game of craps, in which:

- if a 2, 3 or 12 is rolled on the first roll, the shooter loses
- if a 7 or 11 is rolled on the first roll, the shooter wins
- if a 4, 5, 6, 8, 9, or 10 is rolled on the first roll, the dice are rolled again until the either that number or a 7 comes up, in which case the shooter wins or loses, respectively.

In this case there are an infinite number of outcomes in $\mathcal{S}$, some of which are: 2, 3, 7, 11, 12, 4|4, 4|7, 5|5, 5|7, 6|6, 6|7, 8|8, 8|7, 9|9, 9|7, 10|10, 10|7, 4|2|4, 4|3|4, 4|5|4, 4|6|4, ... Possible events include: the shooter wins $\{7, 11, 4|4, 5|5, 6|6, 8|8, \ldots\}$; the shooter loses $\{2, 3, 12, 4|7, \ldots\}$; the dice are thrown exactly once $\{2, 3, 7, 11, 12\}$, etc.

Since an event is a set of outcomes, we can use all of the machinery of set theory, specifically:

- The **complement** A' of a set A, is the set of all outcomes in $\mathcal{S}$ which are *not* in A.
- The **union** $A \cup B$ of two sets A and B, is the set of all outcomes which are in A *or* B, including those which are in both.
- The **intersection** $A \cap B$ is the set of all outcomes which are in both A *and* B.

In the case of coin flips, if the events are $A = \{HHT, HTH, THH\}$ (exactly two heads) and $B = \{HHH, HHT, HTH, HTT\}$ (first flip heads), we

$$A' = \{HHH, HTT, THT, TTH, TTT\}$$
$$A \cup B = \{HHH, HHT, HTH, HTT, THH\}$$
$$A \cap B = \{HHT, HTH\}$$

Another useful definition is that A and B are **disjoint** or **mutually exclusive** events if $A \cap B = \varnothing$.

1.2 Rules of Probability

Having formally defined what we mean by an event, we can proceed to define the probability of that event, which we think of as the chance that it will occur. Devore starts with three axioms

1. For any event A, $P(A) \geq 0$
2. $P(\mathcal{S}) = 1$
3. Given an infinite collection A_1, A_2, A_3, ...of *disjoint* events,

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \qquad (1.1)$$

From there he manages to derive a bunch of other sensible results, such as

1. For any event A, $P(A) \leq 1$
2. $P(\varnothing) = 0$
3. $P(A') = 1 - P(A)$

One useful result concerns the probability of the union of any two events. Since $A \cup B = (A \cap B') \cup (A \cap B) \cup (A' \cap B)$, the union of three disjoint events,

$$P(A \cup B) = P(A \cap B') + P(A \cap B) + P(A' \cap B) \tag{1.2}$$

On the other hand, $A = (A \cap B') \cup (A \cap B)$ and $B = (A \cap B) \cup (A' \cap B)$, so

$$P(A) = P(A \cap B') + P(A \cap B) \tag{1.3a}$$
$$P(B) = P(A \cap B) + P(A' \cap B) \tag{1.3b}$$

which means that

$$P(A) + P(B) = P(A \cap B') + 2P(A \cap B) + P(A' \cap B) = P(A \cup B) + P(A \cap B) \tag{1.4}$$

so

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.5}$$

1.3 Assigning Probabilities

The axioms of probability let us relate the probabilities of different events, but they don't tell us what those probabilities should be in the first place. If we have a way of assigning probabilities to each outcome, and therefore each simple event, then we can use the sum rule for disjoint events to write the probability of any event as the sum of the probabilities of the simple events which make it up. I.e.,

$$P(A) = \sum_{E_i \text{ in } A} P(E_i) \tag{1.6}$$

One possibility is that each outcome, i.e., each simple event might be equally likely. In that case, if there are N outcomes total, the probability of each of the simple events is $P(E_i) = 1/N$ (so that $\sum_{i=1}^{N} P(E_i) = P(\mathcal{S}) = 1$), and in that case

$$P(A) = \sum_{E_i \text{ in } A} \frac{1}{N} = \frac{N(A)}{N} \tag{1.7}$$

where $N(A)$ is the number of outcomes which make up the event A.

Note, however, that one has to consider whether it's appropriate to take all of the outcomes to be equally likely. For instance, in our craps example, we considered each roll, e.g., 2 and 4, to be its own outcome. But you can also consider the rolls of the individual dice, and then the two dice totalling 4 would be a composite event consisting of the outcomes $(1, 3)$, $(2, 2)$, and $(3, 1)$. For a pair of fair dice, the 36 possible outcomes defined by the numbers on the two dice taken in order (suppose one die is green and the other red) are equally likely outcomes.

2 Counting Techniques

2.1 Ordered Sequences

We can come up with 36 as the number of possible results on a pair of fair dice in a couple of ways. We could make a table

	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

which is also useful for counting the number of occurrences of each total. Or we could use something called a tree diagram:

First roll

Second roll

1
2
3

4

5

6
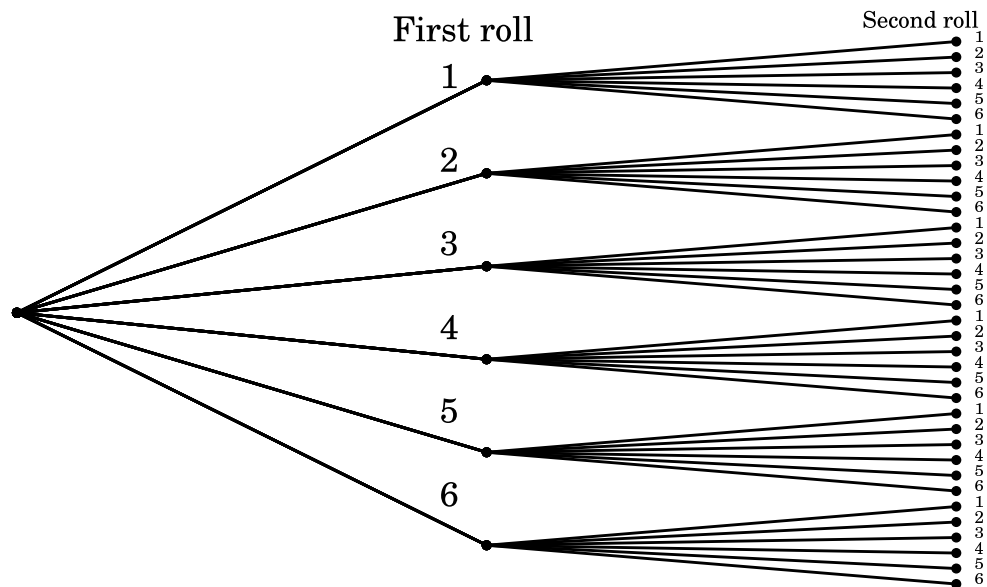
This works well for counting a small number of possible outcomes, but already with 36 outcomes it is becoming unwieldy. So instead of literally counting the possible outcomes, we should calculate how many there will be. In this case, where the outcome is an ordered pair of numbers from 1 to 6, there are 6 possibilities for the first number, and corresponding to each of those there are 6 possibilities for the second number. So the total is $6 \times 6 = 36$.

More generally, if we have an ordered set of k objects, with n_1 possibilities for the first, n_2 for the second, etc, the number of possible ordered k-tuples is $n_1 n_2 \ldots n_k$, which we can also write as

$$\prod_{i=1}^{k} n_k \ . \tag{2.1}$$

2.2   Permutations and Combinations

Consider the probability of getting a poker hand (5 cards out of the 52-card deck) which consists entirely of hearts.[1] Since there are four different suits, you might think the odds are $(1/4)(1/4)(1/4)(1/4)(1/4) = (1/4)^5 = 1/4^5$.

---

[1]This is, hopefully self-apparently, one-quarter of the probability of getting a flush of any kind.

6

However, once a heart has been drawn on the first card, there are only 12 hearts left in the deck out of 51; after two hearts there are 11 out of 50, etc., so the actual odds are

$$P(\heartsuit\heartsuit\heartsuit\heartsuit\heartsuit) = \left(\frac{13}{52}\right)\left(\frac{12}{51}\right)\left(\frac{11}{50}\right)\left(\frac{10}{49}\right)\left(\frac{9}{48}\right) \tag{2.2}$$

This turns out not to be the most effective way to calculate the odds of poker hands, though. (For instance, it's basically impossible to do a card-by-card accounting of the probabability of getting a full house.) Instead we'd like to take the approach of counting the total number of possible five-card hands (outcomes) and then counting up how many fall into a particular category (event). The terms for the quantities we will be interested in are **permutation** and **combination**.

First, let's consider the number of possible sequences of five cards drawn out of a deck of 52. This is the permutation number of permutations of 5 objects out of 52, called $P_{5,52}$. The first card can be any of the 52; the second can be any of the remaining 51; the third can be any of the remaining 50, etc. The number of permutations is

$$P_{5,52} = 52 \times 51 \times 50 \times 49 \times 48 \tag{2.3}$$

In general

$$P_{k,n} = n(n-1)(n-2)\cdots(n-k+1) = \prod_{\ell=0}^{k-1}(n-\ell) \ . \tag{2.4}$$

Now, there is a handy way to write this in terms of the factorial function. Remember that the factorial is defined as

$$n! = n(n-1)(n-1)\cdots(2)(1) = \prod_{\ell=1}^{n}\ell \tag{2.5}$$

with the special case that $0! = 1$. Then we can see that

$$\frac{n!}{(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)\cancel{(n-k)}\cancel{(n-k-1)}\cdots\cancel{(2)}\cancel{(1)}}{\cancel{(n-k)}\cancel{(n-k-1)}\cdots\cancel{(2)}\cancel{(1)}}$$
$$= P_{k,n} \tag{2.6}$$

Note in particular that the number of ways of arranging n items is

$$P_{n,n} = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n! \tag{2.7}$$

Now, when we think about the number of different poker hands, actually we don't consider the cards in a hand to be ordered. So in fact all we care about is the number of ways of choosing 5 objects out of a set of 52, without regard to order. This is the number of **combinations**, which is sometimes written $C_{5,52}$, but which we'll write as $\binom{52}{5}$, pronounced "52 choose 5". When we counted the number of different permutations of 5 cards out of 52, we actually counted each possible hand a bunch of times, once for each of the ways of arranging the cards. There are $P_{5,5} = 5!$ different ways of arranging the five cards of a poker hand, so the number of permutations of 5 cards out of 52 is the number of combinations times the number of permutations of the 5 cards among themselves:

$$P_{5,52} = \binom{52}{5} P_{5,5} \tag{2.8}$$

The factor of $P_{5,5} = 5!$ is the factor by which we overcounted, so we divide by it to get

$$\binom{52}{5} = \frac{P_{5,52}}{P_{5,5}} = \frac{52!}{47!5!} = 2598960 \tag{2.9}$$

or in general

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \tag{2.10}$$

So to return to the question of the odds of getting five hearts, there are $\binom{52}{5}$ different poker hands, and $\binom{13}{5}$ different hands of all hearts (since there are 13 hearts in the deck), which means the probability of the event $A = \heartsuit\heartsuit\heartsuit\heartsuit\heartsuit$ is

$$P(A) = \frac{N(A)}{N} = \frac{\binom{13}{5}}{\binom{52}{5}} = \frac{\frac{13!}{8!5!}}{\frac{52!}{47!5!}} = \frac{13!47!}{8!52!} = \frac{(13)(12)(11)(10)(9)}{(52)(51)(50)(49)(48)} \tag{2.11}$$

which is of course what we calculated before. Numerically, $P(A) \approx 4.95 \times 10^{-4}$, while $1/4^5 \approx 9.77 \times 10^{-4}$. The odds of getting any flush are four times the odds of getting an all heart flush, i.e., 1.98×10^{-3}.

Actually, if we want to calculate the odds of getting a flush, we have over-counted somewhat, since we have also included straight flushes, e.g., 4♡-5♡-6♡-7♡-8♡. If we want to count only hands which are flushes, we need to subtract those. Since aces can count as either high or low, there are ten different all-heart straight flushes, which means the number of different all-heart flushes which are not straight flushes is

$$\binom{13}{5} - 10 = \frac{13!}{8!5!} - 10 = 1287 - 10 = 1277 \tag{2.12}$$

and the probability of getting an all-heart flush is 4.92×10^{-4}, or 1.97×10^{-3} for any flush.

Exercise: work out the number of possible straights and therefore the odds of getting a straight.

Practice Problems

2.5, 2.9, 2.13, 2.17, 2.29, 2.33, 2.43

Thursday 18 March 2010

3 Conditional Probabilities and Tree Diagrams

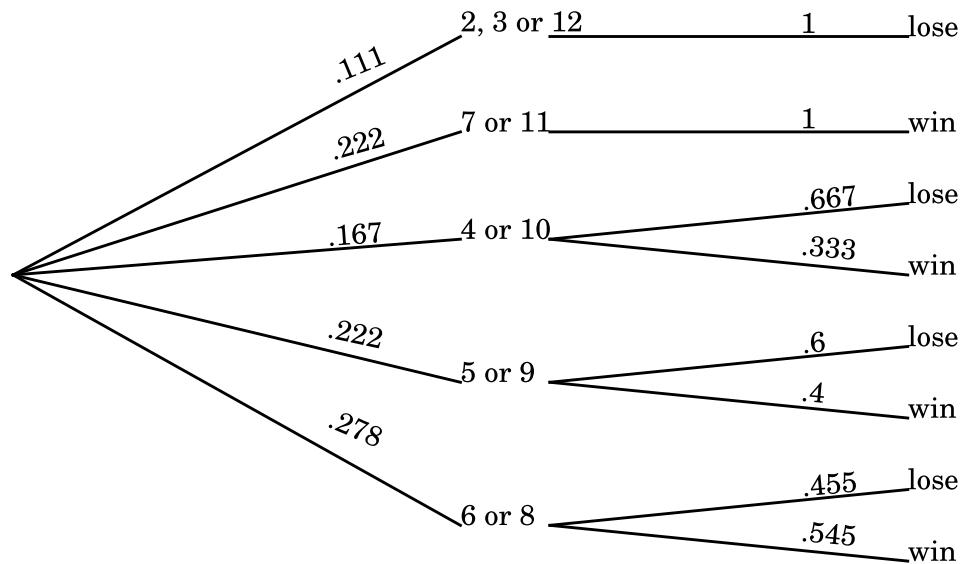## 3.1 Example: Odds of Winning at Craps

Although there are an infinite number of possible outcomes to a craps game, we can still calculate the probability of winning.

First, the sample space can be divided up into mutually exclusive events based on the result of the first roll:

Event	Probability	Result of game
2, 3 or 12 on 1st roll	$\frac{1+2+1}{36} = \frac{4}{36} \approx 11.1\%$	lose
7 or 11 on 1st roll	$\frac{6+2}{36} = \frac{8}{36} \approx 22.2\%$	win
4 or 10 on 1st roll	$\frac{3+3}{36} = \frac{6}{36} \approx 16.7\%$	???
5 or 9 on 1st roll	$\frac{4+4}{36} = \frac{8}{36} \approx 22.2\%$	???
6 or 8 on 1st roll	$\frac{5+5}{36} = \frac{10}{36} \approx 27.8\%$	???

The last three events each contain some outcomes that correspond to winning, and some that correspond to losing. We can figure out the probability of winning if, for example, you roll a 4 initially. Then you will win if another 4 comes up before a 7, and lose if a 7 comes up before a 4. On any given roll, a 7 is twice as likely to come up as a 4 (6/36 vs 3/36), so the odds are $6/9 = 2/3 \approx 66.7\%$ that you will roll a 7 before a 4 and lose. Thus the odds of losing after starting with a 4 are 66.7%, while the odds of winning after starting with a 4 are 33.3%. The same calculation applies if you get a 10 on the first roll. This means that the $6/36 \approx 16.7\%$ probability of rolling a 4 or 10 initially can be divided up into a $4/36 \approx 11.1\%$ probability to start with a 4 or 10 and eventually lose, and a $2/36 \approx 5.6\%$ probability to start with a 4 or 10 and eventually win.

We can summarize this branching of probabilities with a tree diagram:

2, 3 or 12 _____ 1 _____ lose

.111

.222 _____ 7 or 11 _____ 1 _____ win

.667 ___ lose

.167 _____ 4 or 10 _____

.333 ___ win

.222 .6 ___ lose

5 or 9 _____

.278 .4 ___ win

.455 ___ lose

6 or 8 _____

.545 ___ win

The probability of winning given that you've rolled a 4 or 10 initially is an example of a conditional probability. If A is the event "roll a 4 or 10 initially" and B is the event "win the game", we write the conditional probability for event B given that A occurs as $P(B|A)$. We have argued that the probability for both A and B to occur, $P(A \cap B)$, should be the probability of A times the conditional probability of B given A, i.e.,

$$P(A \cap B) = P(B|A)P(A) \tag{3.1}$$

We can use this to fill out a table of probabilities for different sets of outcomes of a craps game, analogous to the tree diagram.

A	$P(A)$	B	$P(B|A)$	$P(A \cap B) = P(B|A)P(A)$
2, 3 or 12 on 1st roll	.111	lose	1	.111
7 or 11 on 1st roll	.222	win	1	.222
4 or 10 on 1st roll	.167	lose	.667	.111
		win	.333	.056
5 or 9 on 1st roll	.222	lose	.6	.133
		win	.4	.089
6 or 8 on 1st roll	.278	lose	.545	.152
		win	.455	.126

Since the rows all describe disjoint events whose union is the sample space $\mathcal{S}$, we can add the probabilities of winning and find that

$$P(\text{win}) \approx .222 + .056 + .089 + .126 \approx .493 \qquad (3.2)$$

and

$$P(\text{lose}) \approx .111 + .111 + .133 + .152 \approx .507 \qquad (3.3)$$

3.2 Definition of Conditional Probability

We've motivated the concept of conditional probability and applied it via (3.1). In fact, from a formal point of view, conditional probability is *defined* as

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \; . \qquad (3.4)$$

We actually used that definition in another context above without realizing it, when we were calculating the probability of rolling a 7 before rolling a 4. We know that $P(7) = 6/36$ and $P(4) = 3/36$ on any given roll. The probability of rolling a 7 given that the game ends on that throw is

$$P(7|7 \cup 4) = \frac{P(7)}{P(7 \cup 4)} = \frac{P(7)}{P(7) + P(4)} = \frac{6/36}{9/36} = \frac{6}{9} \qquad (3.5)$$

We calculated that using the definition of conditional probability.

4   Bayes's Theorem

*S*ome of the arguments in this section are adapted from
`http://yudkowsky.net/rational/bayes`
which gives a nice explanation of Bayes's theorem.

The laws of probability are pretty good at predicting how likely something is to happen given certain underlying circumstances. But often what you really want to know is the opposite: given that some thing happened, what were the circumstances? The classic example of this is a test for a disease.

Suppose that one one-thousandth of the population has a disease. There is a test that can detect the disease, but it has a 2% false positive rate (on average one out of fifty healthy people will test positive) and as 1% false negative rate (on average one out of one hundred sick people will test

negative). The question we ultimately want to answer is: if someone gets a positive test result, what is the probability that they actually have the disease. Note, it is not 98%!

4.1 Approach Considering a Hypothetical Population

The standard treatment of Bayes's Theorem and the Law of Total Probability can be sort of abstract, so it's useful to keep track of what's going on by considering a hypothetical population which tracks the various probabilities. So, assume the probabilities arise from a population of 100,000 individuals. Of those, one one-one-thousandth, or 100, have the disease. The other 99,900 do not. The 2% false positive rate means that of the 99,900 healthy individuals, 2% of them, or 1,998, will test positive. The other 97,902 will test negative. The 1% false negative rate means that of the 100 sick individuals, one will test negative and the other 99 will test positive. So let's collect this into a table:

	Positive	Negative	Total
Sick	99	1	100
Healthy	1,998	97,902	99,900
Total	2,097	97,903	100,000

(As a reminder, if we choose a *sample* of 100,000 individuals out of a larger population, we won't expect to get exactly this number of results, but the 100,000-member population is a useful conceptual construct.)

Translating from numbers in this hypothetical population, we can confirm that it captures the input information:

$$P(\text{sick}) = \frac{100}{100,000} = .001 \tag{4.1a}$$

$$P(\text{positive}|\text{healthy}) = \frac{1,998}{99,900} = .02 \tag{4.1b}$$

$$P(\text{negative}|\text{sick}) = \frac{1}{100} = .01 \tag{4.1c}$$

But now we can also calculate what we want, the conditional probability of being sick given a positive result. That is the fraction of the total number of individuals with positive test results that are in the "sick *and* positive"

category:

$$P(\text{sick}|\text{positive}) = \frac{99}{2,097} \approx .04721 \tag{4.2}$$

or about 4.7%.

Note that we can forego the artificial construct of a 100,000-member hypothetical population. If we divide all the numbers in the table by 100,000, they become probabilities for the corresponding events. For example,

$$P(\text{sick} \cap \text{positive}) = \frac{99}{100,000} = .00099$$

That is the approach of the slightly more axiomatic (and general) method described in the next section.

4.2   Approach Using Axiomatic Probability

The quantity we're looking for (the probability of being sick, given a positive test result) is a conditional probability. To evaluate it, we need to define some events about which we'll discuss the probability. First, consider the events

$$A_1 \equiv \text{"individual has the disease"} \tag{4.3a}$$
$$A_2 \equiv \text{"individual does not have the disease"} \tag{4.3b}$$

These make a mutually exclusive, exhaustive set of events, i.e., $A_1 \cap A_2 = \varnothing$ and $A_1 \cap A_2 = \mathcal{S}$. (We call them A_1 and A_2 because in a more general case there might be more than two events in the mutually exclusive, exhaustive set.) We are told in the statement of the problem that one person in 1000 has the disease, which means that

$$P(A_1) = .001 \tag{4.4a}$$
$$P(A_2) = 1 - P(A_1) = .999 \tag{4.4b}$$
$$\tag{4.4c}$$

Now consider the events associated with the test:

$$B \equiv \text{"individual tests positive"} \tag{4.5a}$$
$$B' \equiv \text{"individual tests negative"} \tag{4.5b}$$

14

(We call them B and B' because we will focus more on B.) The 2% false positive and 1% false negative rates tell us

$$P(B|A_2) = .02 \tag{4.6a}$$
$$P(B'|A_1) = .01 \tag{4.6b}$$

Note that if we want to talk about probabilities involving B, we should use the fact that

$$P(B|A_1) + P(B'|A_1) = 1 \tag{4.7}$$

to state that

$$P(B|A_1) = 1 - P(B'|A_1) = .99 \tag{4.8}$$

Now we can write down the quantity we actually want, using the definition of conditional probability:

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} \tag{4.9}$$

Now, we don't actually have an expression for $P(A_1 \cap B)$ or $P(B)$ yet. The fundamental things we know are

$$P(A_1) = .001 \tag{4.10a}$$
$$P(A_2) = .999 \tag{4.10b}$$
$$P(B|A_2) = .02 \tag{4.10c}$$
$$P(B|A_1) = .99 \tag{4.10d}$$

However, we know how to calculate $P(A_1 \cap B)$ and $P(B)$ from the things we do know. First, to get $P(A_1 \cap B)$, we can notice that we know $P(B|A_1)$ and $P(A_1)$, so we can solve

$$P(B|A_1) = \frac{P(A_1 \cap B)}{P(A_1)} \tag{4.11}$$

for

$$P(A_1 \cap B) = P(B|A_1)P(A_1) = (.99)(.001) = .00099 \tag{4.12}$$

Logically, the probability of having the disease *and* testing positive for it is the probability of having the disease in the first place times the probability of testing positive, given that you have the disease.

15

Since we know $P(B|A_2)$ and $P(A_2)$ we can similarly calculate the probability of not having the disease but testing positive anyway:

$$P(A_2 \cap B) = P(B|A_2)P(A_2) = (.02)(.999) = .01998 \qquad (4.13)$$

But now we have enough information to calcuate $P(B)$, since the overall probability of testing positive for the disease has to be the probability of having the disease and testing positive plus the probability of not having the diesease and testing positive:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) = .00099 + .01998 = .02097 \qquad (4.14)$$

This is an application of the *Law of total probability*, which says, in the more general case of k mutually exclusive, exhaustive events $A_1, \ldots A_k$,

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_k) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k) \qquad (4.15)$$

Now we are ready to calculate

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.0099}{.02097} \approx .04721 \qquad (4.16)$$

So only about 4.7% of people who test positive have the disease. It's a lot more than one in a thousand, but a lot less than 99%.

This is an application of Bayes's theorem, which says

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \qquad (4.17)$$

or, for k mutually exclusive, exhaustive alternatives,

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)} \qquad (4.18)$$

Practice Problems

2.45, 2.59, 2.63, 2.71, 2.105 parts a & b