4.6.3 More-Complicated Calculations Using Conditional Probabilities

Example 4.6.6 ***Using Conditional Probabilities for a Sequence of Events***

Suppose we sample 2 balls at random, one at a time without replacement, from an urn containing 4 black balls and 3 white balls. We want to calculate the probability that the second ball is black. When we come to make the second draw, the chances of drawing a black ball depend on what ball was removed at the first draw because that determines the composition of the balls in the urn. We shall therefore have to use information that comes naturally in the form of conditional probabilities. We use the same notation as in Example 4.6.5, for example, B_2 denotes the event that the second ball sampled is black.

Tree Diagrams

One method that is sometimes used to tackle problems like the one in Example 4.6.6 involves a type of diagram called a *(probability) tree diagram.* These diagrams often provide a convenient way of organizing (and then using) conditional probability information. To aid the discussion, Fig. 4.6.2 gives a tree diagram for the situation in Example 4.6.6. Along the way, we shall state some general rules for constructing and using such trees.

The probability written beside each line segment in the tree is the probability that the right-hand event on the line segment occurs *given* the occurrence of all the events that have appeared along that path so far (reading from left to right). Each time a branching occurs in the tree, we want to cover all eventualities, so the probabilities beside any "fan" of line segments should add to unity.

Because the probability information on a line segment is conditional on what has gone before, the order in which the tree branches should reflect the type of information that is available. In Example 4.6.6 we have unconditional probability information about the first draw so the first set of branches of the tree (represented by B_1 and W_1) concern the first draw. The readily available probability information about the second draw depends on (i.e., is conditional on) what happened at the first draw and thus

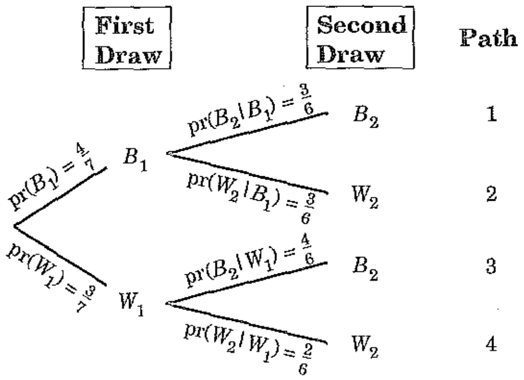

FIGURE 4.6.2 Tree diagram for a sampling problem.

forms the second set of branches. We draw the tree to represent all four possible outcomes: "B_1 and B_2," "B_1 and W_2," "W_1 and B_2," and "W_1 and W_2." These four outcomes (events) are mutually exclusive and give all the possibilities.

Rules for Use

(i) ***Multiply along a path*** to get the probability that all of the events on that path occur. (This uses the multiplication rule for conditional probabilities.)

(ii) ***Add*** the probabilities of ***all whole paths*** in which an event occurs to obtain the probability of that event occurring. (This uses the addition rule for mutually exclusive events.)

EXAMPLE 4.6.6 (cont.) *Using the Rules*

To get the probability of obtaining a black ball on the second draw, the rules tell us to multiply along paths and add whole paths containing a black ball on the second draw (namely, paths 1 and 3). This gives us

$$\text{pr}(B_2) = \frac{4}{7} \times \frac{3}{6} + \frac{3}{7} \times \frac{4}{6} = \frac{4}{7}$$

EXAMPLE 4.6.7 *Tree Diagram and Two-Way Table Methods*

We revisit the data in Example 4.6.4. A 1992 news report stated that 11% of Israel's Jewish population and 52% of its Arabic citizens lived below the poverty line. Arabic citizens were reported to make up 14% of the population of Israel. We shall assume that these two groups account for the whole population of Israel so that 86% of the population is Jewish. We shall determine (i) proportion of the Israel's population living below the poverty line, and (ii) the proportion of poor people in Israel who were Arabic. In probability notation, we can write the information we have been given as $\text{pr}(\textit{Poor} \mid \textit{Jewish}) = 0.11$, $\text{pr}(\textit{Poor} \mid \textit{Arabic}) = 0.52$, $\text{pr}(\textit{Arabic}) = 0.14$ and $\text{pr}(\textit{Jewish}) = 0.86$.

There are two factors at work here: ethnic group and whether or not someone is poor. We could use a tree to work on this problem splitting first on ethnicity, because we have unconditional information about this, and then on poverty because our information about poverty is conditional on what ethnic group is being discussed. This has been done in Fig. 4.6.3.

The event "being poor" corresponds to paths 1 and 3. Our rules tell us to obtain $\text{pr}(\textit{Poor})$ by multiplying along paths and adding whole paths containing this event. Thus, $\text{pr}(\textit{Poor}) = 0.14 \times 0.52 + 0.86 \times 0.11 = 0.1674$. This tells us that approximately 17% of the population of Israel lives below the poverty line. We can now find the proportion of poor people who are Arabic by

$$\text{pr}(\textit{Arabic} \mid \textit{Poor}) = \frac{\text{pr}(\textit{Poor} \text{ and } \textit{Arabic})}{\text{pr}(\textit{Poor})} = \frac{0.14 \times 0.52}{0.1674} = 0.4349$$

We see that almost half (43%) of Israel's poor are Arabic.

This approach may look reasonably simple—once we have laid out the tree for you! We have found, however, that many of our students are better able to solve problems using the table method to follow. Our two factors, ethnicity and poverty, become

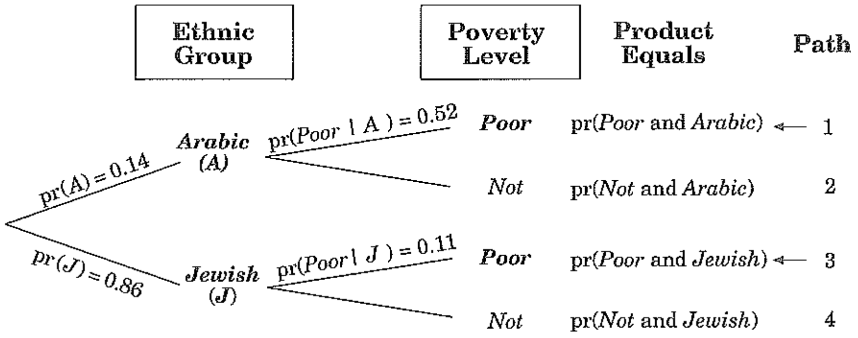

FIGURE 4.6.3 Poverty in Israel.

pr(*Poor* and *Arabic*) = pr(*Poor* | *Arabic*) × pr(*Arabic*) [= 52% of 14%]

pr(*Poor* and *Jewish*) = pr(*Poor* | *Jewish*) × pr(*Jewish*) [= 11% of 86%]

		Ethnicity		
		Arabic	Jewish	Total
Poverty	Poor	.52 × .14	.11 × .86	?
	Not poor	?	?	?
	Total	.14	.86	1.00

pr(*Arabic*) = .14 pr(*Jewish*) = .86

FIGURE 4.6.4 Proportions by ethnicity and poverty.

the two dimensions of a two-way table. Our information that 14% of the population are Arab and 86% are Jewish belong in the column totals of the table. The entries inside the two-way table of proportions need to be of the form pr(A and B), not conditional information. Thus our conditional information, for example, pr(*Poor* | *Arabic*) = 0.52, cannot be inserted as it stands. However, we can construct the right type of proportions using the multiplication rule pr(A and B) = pr(B | A) pr(A), as shown in Fig. 4.6.4.

We then fill in all the unknown entries in the table, using the simple idea that sums across rows have to give row totals and sums down columns must produce the column totals, as follows.

.0728	.0946	?
?	?	?
.14	.86	1.00

→

.0728	.0946	.1674
?	?	?
.14	.86	1.00

→

.0728	.0946	.1674
.0672	.7654	.8326
.14	.86	1.00

We got from the second to the third of the small tables above by making the columns add up. The third small table is our completed table, which we present with all the labels attached as Table 4.6.3. Many probabilities can be read off Table 4.6.3, including pr(*Poor*) = 0.1674. We can obtain any of the conditional probabilities in the normal way including

$$\text{pr}(\textit{Jewish} \mid \textit{Poor}) = \frac{\text{pr}(\textit{Poor} \text{ and } \textit{Jewish})}{\text{pr}(\textit{Poor})} = \frac{0.0946}{0.1674} = 0.5651$$

TABLE 4.6.3 Proportions by Ethnicity and Poverty

		Ethnicity		
		Arabic	Jewish	Total
Poverty	Poor	.0728	.0946	.1674
	Not poor	.0672	.7654	.8326
	Total	.14	.86	1.00

We note that in pr(*Jewish* | *Poor*) the order of the conditioning is reversed from the order in the information given to us, which was of the form pr(*Poor* | *Jewish*). The table has allowed us to reverse the order of the conditioning.[14]

> Many problems involving ***the reversal of order of conditional probabilities*** can be solved by constructing two-way tables.

EXAMPLE 4.6.8 *Constructing a Two-Way Table*

From a 1990 study issued by the National Academy of Sciences in the United States it was found that, of American women using contraception, 38% are sterilized, 32% use oral contraceptives, 24% use barrier methods (diaphragm, condom, cervical caps), 3% use IUDs, and 3% rely on spermicides (foams, creams, jellies). If we define the failure rates of a method as the percentage of those who become pregnant during a year of use of the method, then the failure rates for each of these methods are approximately as follows: sterilization 0%, the oral contraceptive pill 5%, barrier methods 14%, IUDs 6%, and spermicides 26%. One question we would like to answer is, What percentage of women using contraception experience an unwanted pregnancy over the course of a year? Another is, If we look only at women who experienced contraceptive failure, what proportions were using each type of contraceptive method?

There are two factors at work in this problem as well. One is *Method*—the type of contraceptive method the woman was using. The other is *Outcome*—whether or not the woman experienced contraceptive failure. This suggests constructing a two-way table with dimensions *Method* and *Outcome*, as in Fig. 4.6.5. We are given information about the proportions of women using each method, pr(*Sterilized*) = 0.38, pr(*Oral*) = 0.32, ... This information can be inserted directly into the column totals, as in Fig. 4.6.5. Our other information is about the proportion of women experiencing failure conditional on the type of contraceptive used, for example, we see pr(*Fail* | *Sterilized*) = 0, pr(*Fail* | *Oral*) = 0.05, pr(*Fail* | *Barrier*) = 0.14, and so on. This cannot be inserted directly because the interior entries in the table must be of the form pr(*A* and *B*) rather than conditional probabilities. Once more, we construct these from the conditional probabilities using the multiplication rule, as shown in Fig. 4.6.5.

[14]The two-way table has avoided the use of the so-called Bayes' theorem, which is a more traditional way of solving these problems in other textbooks.

pr(*Failed* and *Oral*) =
pr(*Failed* | *Oral*) × pr(*Oral*)
[= 5% of 32%]

pr(*Failed* and *IUD*) =
pr(*Failed* | *IUD*) × pr(*IUD*)
[= 6% of 3%]

		Method					
		Steril.	Oral	Barrier	IUD	Sperm.	Total
Outcome	Failed	0 × .38	.05 × .32	.14 × .24	.06 × .03	.26 × .03	?
	Didn't	?	?	?	?	?	?
	Total	.38	.32	.24	.03	.03	1.00

pr(*Steril.*) = .38 pr(*Barrier*) = .24 pr(*IUD*) = .03

FIGURE 4.6.5 Proportions by outcome and method.

TABLE 4.6.4 Table Constructed from the Data in Example 4.6.8

		Method					
		Steril.	**Oral**	**Barrier**	**IUD**	**Sperm.**	**Total**
Outcome	Failed	0	.0160	.0336	.0018	.0078	.0592
	Didn't	.3800	.3040	.2064	.0282	.0222	.9408
	Total	.3800	.3200	.2400	.0300	.0300	1.0000

Knowing that the table should add up going across rows and down columns enables us to fill in the missing entries and obtain Table 4.6.4.

From the completed table, we are able to read off pr(*Failed*) = 0.0592, or approximately 6% of the women sampled had experienced contraceptive failure. Conditional probabilities such as pr(*Barrier* | *Failure*) are easily constructed from the information in the table:

$$\text{pr}(\textit{Barrier} \mid \textit{Failure}) = \frac{\text{pr}(\textit{Barrier} \text{ and } \textit{Failure})}{\text{pr}(\textit{Failure})} = \frac{0.0336}{0.0592} = 0.568$$

This tells us that, of the women experiencing contraceptive failure, 57% were using IUDs. This sort of information is often quoted in the media. If you read those words, you would tend to think, "IUDs must be a particularly unreliable form of contraception." However, pr(*IUD* | *Failure*) is the wrong probability for deciding which method to use. The relevant probability is pr(*Failure* | *IUD*), that is, the failure rate among those using the method. For IUDs this is 6%, which is nearly as good as the 5% failure rate for oral contraceptives and much better than barrier methods or spermicides.

EXERCISES FOR SECTION 4.6.3

Solve the following problems by setting up an appropriate two-way table.

1. In New Zealand, 3.24% of Europeans and 1.77% of Maori have type AB blood. A blood bank in a district where the population is 85% European and 15% Maori wants to know how much AB blood to stock. What percentage of people in the district have AB blood? What percentage of the people in the AB blood group are Maori?

2. The chances of a child being left-handed are 1 in 2 if both parents are left-handed, 1 in 6 if one parent is left-handed, and 1 in 16 if neither parent is left-handed (*New Zealand Herald*, 5 January 1991). Suppose that, of couples having children, in 2% both father and mother are left-handed, in 20% one is left-handed, and in the rest neither is left-handed. What is the probability of a randomly chosen child being left-handed? What is the probability that neither parent of a left-handed child is left-handed?

3. University of Florida sociologist Michael Radelet believed that if you killed a white person in Florida, the chances of getting the death penalty were three times greater than if you had killed a black person (*Gainesville Sun*, 20 October 1986). In a study, Radelet classified 326 murderers by race of the victim and type of sentence given to the murderer. He found that 36 of the convicted murderers received the death sentence. Of this group, 30 had murdered a white person, whereas 184 of the group that did not receive the death sentence had murdered a white person. If a victim from this study was white, what is the probability that the murderer of this victim received the death sentence? Do you agree with Radelet?

QUIZ ON SECTION 4.6

1. In pr($A \mid B$), how should the symbol "|" be read? (Section 4.6.1)
2. Give an example where A and B are two events with pr(A) $\neq 0$ but pr($A \mid B$) $= 0$.
3. If event A always occurs when B occurs, what can you say about pr($A \mid B$)?
4. When drawing a probability tree for a particular problem, how do you know what events to use for the first fan of branches and which to use for the subsequent fans?
   What probability do you use to label a line segment?
   How do you find the probability that all events along a given branch occur?
   How do you find the probability that a particular event occurs? (Section 4.6.3)
5. What results do you use for two-way tables to fill in unknown entries?

CASE STUDY 4.6.1 *Testing for AIDS*

It is well known that AIDS is one of the most important public health problems facing the world today. The sense of extreme crisis has dulled in western countries as the 1990s end, but AIDS is rampant in parts of Africa. AIDS is believed to be caused by the human immunodeficiency virus (HIV), but many years can elapse between HIV infection and the development of AIDS. This case study is based on the situation in the early '90s when there were still widespread demands for screening whole populations for HIV infection. In 1990 the World Health Organization (WHO) projected between 25 and 30 million cases of HIV infection worldwide by the year 2000. The United States, where 200,000 AIDS cases had been reported by mid-1992, was the worst-affected western country, largely because the epidemic began earlier in the United States. By 1990 WHO estimated that one in every 75 males and one in every 700 females in the United States was infected with HIV.

The enzyme-linked immunosorbent assay (ELISA) test was the main test used to screen blood samples for antibodies to the HIV virus (rather than the virus itself). It gives a measured mean absorbance ratio for HIV (previously called HTLV) antibodies. Table 4.6.5 gives the absorbance ratio values for 297 healthy blood donors and 88 HIV patients. Healthy donors tend to give low ratios, but some are quite high, partly because the test also responds to some other types of antibody, such as human leucocyte antigen or HLA (Gastwirth [1987, p. 220]). HIV patients tend to have high ratios, but a few give lower values because they have not been able to mount a strong immune reaction.

To use this test in practice, we need a cutoff value so that those who fall below the value are deemed to have tested negatively and those above to have tested positively. Any such cutoff will involve misclassifying some people without HIV as having a positive HIV test (which will be a huge emotional shock), and some people with HIV as having a negative HIV test (with consequences to their own health, the health of people about them, the integrity of the

TABLE 4.6.5 Number of Individuals Having a Given Mean Absorbance Ratio (MAR) in the ELISA for HIV Antibodies

MAR	Healthy donor	HIV patients
< 2	202 (275)	0 (2)
2– 2.99	73	2
3– 3.99	15	7
4– 4.99	3	7
5– 5.99	2	15
6–11.99	2	36
12+	0	21
Total	297	88

Adapted from Weiss et al. [1985].

blood bank, etc.). Using a cutoff ratio of 3 we find that of the healthy people[15] in Table 4.6.5, 275/297 = 0.926 test negatively (22 false positives) and of HIV patients 86/88 = 0.98 test positively (2 false negatives). It should be noted that the false-negative rate may be an undercount.[16] Better results than these have been obtained with the multiple use of ELISA (Gastwirth [1987, p. 236]) and with modern commercial versions of the test. The proportions given above are only rough estimates from small samples. Nevertheless, making use of the numerical equivalence between proportions of a population and probabilities for a randomly chosen individual (Section 4.4.5), we shall use these proportions as if they were true probabilities. Hence

$$\text{pr}(\textit{Positive} \mid \textit{HIV}) = 0.98$$

and, rounding off to two decimal places (as the information is very approximate),

$$\text{pr}(\textit{Negative} \mid \textit{Not HIV}) = 0.93$$

We shall consider the effect of screening the whole U.S. population for AIDS in 1991. At that time, the proportion of Americans infected with HIV was about 1%. We are interested in the proportion of Americans who would test positively and the proportion of those testing positively who would actually have AIDS.

There are two factors of interest in this problem. First is *disease status*—a person either has HIV or does not. Second is *test result*—the person's test result is either positive or negative. We shall construct a two-way table in the usual way to form Fig. 4.6.6. We know that

$$\text{pr}(\textit{Positive} \mid \textit{HIV}) = 0.98$$

$$\text{pr}(\textit{Negative} \mid \textit{Not HIV}) = 0.93$$

$$\text{pr}(\textit{HIV}) = 0.01$$

and thus

$$\text{pr}(\textit{Not HIV}) = 0.99$$

We shall place this information into the table in the usual way, recalling that entries in the interior of the table have to be of the form pr(*A* and *B*). They cannot be conditional.

We now complete the table using the simple idea that the rows and columns of the table must add up to give the

pr(*HIV* and *Positive*) = pr(*Positive* | *HIV*) × pr(*HIV*) [= 98% of 1%]

pr(*Not HIV* and *Negative*) = pr(*Negative* | *Not HIV*) × pr(*Not HIV*) [= 93% of 99%]

		Test result: Positive	Test result: Negative	Total	
Disease status	HIV	.98 × .01	?	.01	← pr(*HIV*) = .01
	Not HIV	?	.93 × .99	.99	← pr(*Not HIV*) = .99
	Total	?	?	1.00	

FIGURE 4.6.6 Putting HIV information into the table.

[15] In the medical and biostatistical literatures, the probability of correctly diagnosing a sick individual as "sick" is called the *sensitivity* of a test, while the probability of correctly diagnosing that a healthy individual does not have the condition of interest is called the *specificity* of that test.

[16] It appears that the virus takes 6 to 12 weeks to provoke antibody production (*Time*, 2 March 1987, p. 44). Also, *Time* (12 June 1989) reports cases of infected men who had not produced antibodies for up to three years.

totals as follows:

.0098	?	.01
?	.9207	.99
?	?	1.00

→

.0098	.0002	.01
.0693	.9207	.99
?	?	1.00

→

.0098	.0002	.01
.0693	.9207	.99
.0791	.9209	1.00

This gives us Table 4.6.6 as our completed table.

We can read pr(*Positive*) = 0.0791 off Table 4.6.6. This is a surprising result. Although only 1% of the population have HIV, 8% would test positively. The majority of the people in any sample who tested positively would not in fact have HIV.[17] This is despite the fact that the test seems reasonably good. After all, it correctly classifies 98% of people with HIV as having HIV and 93% of people without HIV are classified as being HIV free.

TABLE 4.6.6 Proportions by Disease Status and Test Result

		Test result		
		Positive	Negative	Total
Disease status	HIV	.0098	.0002	.01
	Not HIV	.0693	.9207	.99
	Total	.0791	.9209	1.00

So what proportion of people testing positively on ELISA would actually have HIV? We want pr(*HIV* | *Positive*). The conditioning here is in the reverse order to that in the original information supplied to us. We can get everything we need from Table 4.6.6, however.

$$\text{pr}(HIV \mid Positive) = \frac{\text{pr}(HIV \text{ and } Positive)}{\text{pr}(Positive)} = \frac{0.0098}{0.0791} = 0.124$$

As the previous figures suggested, if the whole U.S. population had been screened, only 1 person in 8 (12.4%) who tested positively would have had HIV. The other 7 out of every 8 would be false positives. The situation in other western countries would have been even more extreme, as Table 4.6.7 shows.

The values of pr(*HIV* | *Positive*) given for each of the countries were calculated exactly as above. The only thing that changed from country to country was the value used for pr(*HIV*), the proportion of the population with HIV. [Note that the American figures have changed slightly due to the use of pr(*HIV*) = .00864 rather than the rounded value of 0.01 used in the detailed calculations.]

The most extreme case in Table 4.6.7 is Ireland. If the Irish government had decided to screen the total population of 3.6 million people for HIV in 1990, from the figures above, roughly 250,000 (7%) would have tested positively, and of these, only about 1250 (0.5% of the positives) would

TABLE 4.6.7 Proportions Infected with HIV

Country	No. AIDS[a] cases	Population[b] (millions)	pr(*HIV*)[c]	pr(*HIV* | *Positive*)
United States	218,301	252.7	0.00864	0.109
Canada	6,116	26.7	0.00229	0.031
Australia	3,238	16.8	0.00193	0.026
New Zealand	323	3.4	0.00095	0.013
United Kingdom	5,451	57.3	0.00095	0.013
Ireland	142	3.6	0.00039	0.005

[a]Source: AIDS—New Zealand, November 1992.

[b]1991 estimates, except for Ireland, for which May 1990 figures are given.

[c]Proportion of population infected by HIV. These are very rough. We have assumed that the proportion of HIV-infected people is 10 times larger than the proportion of AIDS cases. This is the approximate relationship between the number of U.S. cases and the U.S. Centers for Disease Control's estimate of the number of HIV-infected Americans in 1990.

[17]People without HIV who test positively are called *false positives*.

have HIV. How do we tell these 1250 people apart from the rest of the 250,000? In the case of HIV there was another more expensive and more specific test, called the Western blot test, that could be used.[18] Thus any screening program would have to include funding for both ELISA tests for everyone and Western blot tests for a quarter of a million people.[19]

Although the value of pr(*HIV* | *Positive*) in Table 4.6.7 varies with the proportion of people with HIV in the population to some extent, all of the entries in the table are small. We don't want to leave you with the reverse misapprehension that pr(*HIV* | *Positive*) is always small. Among intravenous drug users in New York in 1988, it was estimated that 86% had HIV (*New Zealand Herald*, 17 November 1988). Using a value of pr(*HIV*) = 0.86 in the calculations produces pr(*Positive*) = 0.853 and pr(*HIV* | *Positive*) = 0.988. If all New York drug addicts had been screened, almost every person testing positively (98.8%) would have had HIV.

This sort of "good-but-imperfect-test" situation is widespread. It applies to large numbers of medical screening procedures (diabetes, cervical cancer, breast cancer, etc.).[20] It applies to polygraph lie detector tests (some people who are not lying show the physiological symptoms interpreted as a sign of lying, while some people who are lying do not). It applies to psychological and intellectual tests performed to judge the suitability of job applicants (some people who are capable of doing the job well will fail the tests, while some who are not will pass the tests). It can also apply to the testing of urine or blood samples to detect drug use. The type of problem we see in Table 4.6.7, where the majority of those who would test positive would be false positives, is very common in screening for relatively rare conditions. Similar behavior could be expected in testing for drug use among a population in which drug use is rare or using lie detector tests on a group of people in which the vast majority told the truth. An alternative strategy, as indicated by the results for New York drug addicts, is to try to identify high-risk subpopulations and screen only those. With medical screening, particularly in an area as sensitive as AIDS, this can be political dynamite.

So far, we have been using pr(*HIV* | *Positive*) to think about the proportion of those testing positive in a screened population who actually have HIV. But what does pr(*HIV* | *Positive*) mean for an individual?

Let's get personal and imagine that you have just tested positive. Clearly, this would be a major trauma for you. *Time* (2 March 1988) quoted a health professional as saying, "The test tends to rip people's lives apart." *The Economist* (4 July 1992) told a story of a young American who committed suicide on learning that he had tested positive for HIV: "He believed his chance of carrying the virus was 96%. It was 10%."[21] What is pr(*HIV* | *Positive*) for me, that is, what is the probability that *I* have HIV given that *I* have just tested positive on an ELISA test?

We have to think in terms of being a random representative of some population. We saw earlier that the value of pr(*HIV* | *Positive*) depends critically on the value of pr(*HIV*) for the population from which the individual is sampled. None of us can usefully be thought of as a randomly selected individual from our own country as far as HIV is concerned because we know that HIV is much more prevalent in some sections of the population than others. To obtain a value of pr(*HIV* | *Positive*) for yourself, a value for pr(*HIV*) is required that gives the proportion of people who have HIV *among people as much as possible like yourself* with respect to the known risk factors for AIDS. If you are a New York drug addict who shares needles, a positive ELISA test is fairly conclusive. If you have always lived in a monogamous sexual relationship, believe your partner to have done the same, don't share needles, and didn't have a blood transfusion prior to the testing of the blood supply, a positive ELISA test is almost certainly a false positive.

4.7 STATISTICAL INDEPENDENCE

4.7.1 Two Events

We have seen (e.g., Example 4.6.3) that our assessment of the chances that an event occurs can change drastically depending on the information we have about other

[18]In medical terms (see footnote 15) the Western blot is more specific but not as sensitive as ELISA.

[19]Such testing is not cheap! The state of Illinois introduced screening as a condition for a marriage license in 1988. In the first 11 months 150,000 people were screened at a cost of $5.5 million (23 were infected). Many other states now do similar screening.

[20]A screening test, however, is designed to identify a group at increased risk of a condition.

[21]It is surprising that someone was given the results of a positive result on a single test. In New Zealand, people are not told that they have tested positive unless they have also tested positive on a second ELISA test and on a Western blot test.

TABLE 4.7.1 Blood Type Data

.08 × .81 .92 × .81 pr(RH+) = .81

	K+	K-	Total
Rh+	?	?	.81
Rh-	?	?	.19
Total	.08	.92	1.00

.08 × .19 pr(K+) = .08 .92 × .19

→

	K+	K-	Total
Rh+	.0648	.7452	.81
Rh-	.0152	.1748	.19
Total	.08	.92	1.00

probability that a randomly selected person matches this type? The answer is rea from Table 4.7.1, namely 0.0152, which is quite small. Suppose further that one o the suspects has this blood type. Either the suspect is innocent and a random even with probability 0.0152 has occurred, or the suspect is guilty. What do you think?

EXERCISES FOR SECTION 4.7.1

1. According to a study on 3433 women conducted by the Alan Guttmacher Institute in th United States (*Globe and Mail*, 7 August 1989), 6% of women on the contraceptive pill ca expect to become pregnant in the first year compared with 14% of women who do not us the pill but whose partners use condoms. What are the chances of the woman becomin pregnant in the first year if she is on the pill *and* he uses condoms? (Assume independenc Is this a reasonable assumption?)
2. White North Americans in California have blood phenotypes A, B, O, and AB with prob bilities 0.41, 0.11, 0.45, and 0.03, respectively. If two whites are chosen at random, what i the probability that they have the same phenotypes? Why may you assume that the two ar independent?

4.7.2 Positive and Negative Association

In humans, independence of characteristics, as in Example 4.7.1, tends to be the ex ception rather than the rule. Some things we know tend to go together, for exampl blond hair and blue eyes. Someone with blond hair is much more likely to have blu eyes than someone with brown or black hair. We say the events "having blond hair and "having blue eyes" are positively associated. Suppose we look at the populatio in which 30% have blond hair and 25% have blue eyes. If we assumed independenc we would say that the proportion with both is

$$\text{pr}(\textit{blond hair} \text{ and } \textit{blue eyes}) = \text{pr}(\textit{blond hair})\,\text{pr}(\textit{blue eyes}) = 0.3 \times 0.25$$
$$= 0.075$$

Since these events are not independent, we should have used

$$\text{pr}(\textit{blond hair} \text{ and } \textit{blue eyes}) = \text{pr}(\textit{blond hair})\,\text{pr}(\textit{blue eyes} \mid \textit{blond hair})$$
$$= 0.3 \times ?$$

Among blond-haired people the proportion with blue eyes is high, probably muc closer to 80% than 25%. The product 0.3 × 0.8 is then much larger than 0.075. Assun