2.16 PB & J. Suppose 80% of people like peanut butter, 89% like jelly, and 78% like both. Given that a randomly sampled person likes peanut butter, what's the probability that he also likes jelly?

Let's translate "Given that a randomly sampled person likes peanut butter, what's the probability that he also likes jelly?" into probability notation:

If we set P(PB) to be probability of liking peanut butter, and P(J) to be the probability of liking jelly, then "probability of liking jelly given liking peanut butter" would be P(J | PB).

What is P (J | PB)? We know that conditional probability tells us that

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$ (see p. 83 of the text)

$$\text{So, } P(J \mid PB) = \frac{P(J \text{ and } PB)}{P(PB)}$$

What's P(J and PB)? It's 78%, or 0.78. P(PB) is 80%, or 0.80. Dividing .78 by .80, we get **0.975, or 97.5%.** We can do a "sanity check" on this answer by asking ourselves, "does most everyone who likes peanut butter also like jelly?" Since the question tells us that 80% of people like peanut butter and 78% like both, we can say that there's only 2% of the population who like peanut butter without simultaneously liking jelly. Our answer makes sense.

2.18 Weight and health coverage, Part II. Exercise 2.14 introduced a contingency table summarizing the relationship between weight status, which is determined based on body mass index (BMI), and health coverage for a sample of 428,638 Americans. In the table below, the counts have been replaced by relative frequencies (probability estimates).

		Weight Status			
		Neither overweight nor obese (BMI < 25)	Overweight (25 ≤ BMI < 30)	Obese (BMI ≥ 30)	Total
Health Coverage	Yes	0.3145	0.3306	0.2503	0.8954
	No	0.0352	0.0358	0.0336	0.1046
	Total	0.3497	0.3664	0.2839	1.0000

(a) What is the probability that a randomly chosen individual is obese?

I look at the probability in the Total of the Obese column (since I'm not limiting by health insurance coverage status) to discover that the probability that a randomly chosen individual is obese is **0.2839, or 28.39%.**

(b) What is the probability that a randomly chosen individual is obese given that he has health coverage?

Again, since we are talking about conditional probability, we want to divide "both obese and health coverage = yes" by "health coverage = yes". Since the probability of having health coverage is 0.8954 and the probability of both being obese and having health coverage is 0.2503, we divide the latter by the former to get **0.2795.** Since we have a chart to look at, the conditional probability is more intuitive, since we're dividing a part by the whole (the "obese" part of the "Yes, has health insurance" row by the whole or "total" of that row).

(c) What is the probability that a randomly chosen individual is obese given that he doesn't have health coverage?

Similarly to part (b) above, we're dividing part of the no-insurance group 0.0336, the obese part) by the whole no-insurance group (0.1046, or the total probability of not having insurance). The probability that a randomly chosen individual is obese given that he doesn't have health coverage is 0.0336 / 0.1046, or **0.3212**.

(d) Do being overweight and having health coverage appear to be independent?

If being overweight and having health coverage were independent, we'd expect to see that the probability of being overweight would be the same whether or not you had health insurance, and that the probability of having health insurance coverage would be the same regardless of whether or not you were overweight. In other words, knowing your status in one aspect would give us no foresight into your status in the other.

Let's presume here that the term "overweight" in the question means "merely overweight" and excludes obesity, so that we're dealing with just the middle range of BMI's for our chart. (Aside: We can already tell that **obesity** and health coverage are <u>not</u> independent, since the answers to (b) and (c) above were different and indicate a slightly higher possibility of obesity if one is without health insurance coverage.)

We could solve this problem by repeating (b) and (c) calculations for "overweight" instead of "obese" to see if the probabilities are different, or we could go back to the mathematical formula for conditional probability, namely, that

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

We know that A and B are independent iff P(A and B) = P(A) * P(B). Therefore, if we take as a a premise that A and B are independent, the following is true:

$$P(A \mid B) = \frac{P(A) \times P(B)}{P(B)}$$

which means $P(A \mid B) = \dfrac{P(A) \times \cancel{P(B)}}{\cancel{P(B)}}$

and therefore P(A | B) = P(A).

Similarly, we can reverse the A and B and discover that for independent events, P(B | A) = P(B)

So, if "being overweight" and "having health insurance" are independent, the following should be true:

P(being overweight | having health insurance) = P (being overweight)

P(having health insurance | being overweight) = P(having health insurance)

We plug in the values and check:

$$P(\text{being overweight} \mid \text{having health insurance}) \stackrel{?}{=} P(\text{being overweight})$$

$$\frac{P(\text{being overweight and having health insurance})}{P(\text{having health insurance})} \stackrel{?}{=} P(\text{being overweight})$$

$$\frac{0.3306}{0.8954} \stackrel{?}{=} 0.3664$$

$.3692 \stackrel{?}{=} 0.3664$ — Not quite, but very close, perhaps due to significant digit or rounding error or simply due to sample variations from the population

$$P(\text{having health insurance} \mid \text{being overweight}) \stackrel{?}{=} P(\text{having health insurance})$$

$$\frac{\text{P(having health insurance and being overweight)}}{\text{P(being overweight)}} \stackrel{?}{=} \text{P (having health insurance)}$$

$$\frac{0.3306}{0.3664} \stackrel{?}{=} 0.8954$$

$$.3692 \stackrel{?}{=} 0.8954$$

$$0.8954 \stackrel{?}{=} 0.8954 \text{ — Yes, identical within four decimal places.}$$

Without delving into inferential statistics and confidence intervals, the values are sufficiently close to indicate that yes, a BMI status of "overweight" and the fact of having health insurance coverage **seem independent.**

2.20 Assortative mating. Assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Researchers studying this topic collected data on eye colors of 204 Scandinavian men and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.

		Partner (female)			
		Blue	Brown	Green	Total
Self (male)	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	**108**	**55**	**41**	**204**

(a) What is the probability that a randomly chosen male respondent or his partner has blue eyes?

We know that the probability of both having blue eyes is the sum of the probability of each one having blue eyes minus the probability of both having blue eyes (so as to avoid overcounting). This is 114/204 + 108/204 - 78/204, or **0.71.**

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?

We have a given here, so we need to use conditional probability:

$$P(\text{blue-eyed partner} \mid \text{blue-eyed self}) = \frac{P(\text{blue-eyed partner and blue-eyed self})}{P(\text{blue-eyed self})}$$

$$= \frac{78/204}{114/204}$$

$$= \frac{78}{114}$$

(which would be another way to start, just using counts of part of blue-eyed male's partner eye color / whole blue-eyed male's partner eye color)

= **0.68**

(c) What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes? What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?

P (blue-eyed partner | brown-eyed self) = 19/54 or **0.35**

P (blue-eyed partner | green-eyed self) = 11/36 or **0.30**

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

The eye colors of male respondents and their partners are **NOT independent**. We know this because independence means that knowing the outcome of one event (say, the male eye color) does not give us any new information about the other event (in our case, the eye color of the female partner). However, in our case, knowing that a man has blue eyes tells us rather a lot about the eye color of the female partner: it is well over half likely that the parter eye color is blue. On the other hand, if I know that a male's eye color is brown or green, it is far less likely that the partner's eye color is blue. The differences are great enough (without getting into inferential statistics) for us to feel confident that these probability differences are not due to noise in the sample but instead indicate that the two events are not independent.

2.26 Twins. About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex – half are males and the other half are females. One-quarter of fraternal twins are both male, one-quarter both female, and one-half are mixes: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

Let's construct a tree diagram, which will help us visualize the potential outcomes and weigh the probabilities.

I have twins. Are they…

- Fraternal/Identical?
  - Fraternal 0.70
    - Both Girls 0.25
    - Not Both Girls 0.75
  - Identical 0.30
    - Both Girls 0.50
    - Not Both Girls 0.50

P (Identical | Both Girls) = P (Identical and Both Girls) / P(Both Girls)

P(Both Girls) = the sum of the two branches that end with "Both Girls"

$= (0.70)(0.25) + (0.30)(0.50)$

$= 0.325$

P (Identical and Both Girls) = bottom "Both Girls" branch

$= (0.30)(0.50)$

$= 0.15$

P(Identical | Both Girls) = 0.15 / 0.325 = 0.46

There is a 46% chance that if my twins are both girls, they are identical.