Quantifying Uncertainty

Chapter 13

13.1 Show from first principles that $P(a \mid b \wedge a) = 1$.

The “first principles” needed here are the definition of conditional probability, $P(X \mid Y) = P(X \wedge Y) / P(Y)$, and the definitions of the logical connectives. It is not enough to say that if $B \wedge A$ is “given” then A must be true! From the definition of conditional probability, and the fact that $A \wedge A \Leftrightarrow A$ and that conjunction is commutative and associative, we have

$$P(a \mid b \wedge a) = P(a \wedge b \wedge a) / P(b \wedge a) = P(b \wedge (a \wedge a)) / P(b \wedge a) = P(b \wedge a) / P(b \wedge a) = 1$$

13.3 For each of the following statements, either prove it is true or give a counterexample.

a. If $P(a \mid b, c) = P(b \mid a, c)$, then $P(a \mid c) = P(b \mid c)$

b. If $P(a \mid b, c) = P(a)$, then $P(b \mid c) = P(b)$

c. If $P(a \mid b) = P(a)$, then $P(a \mid b, c) = P(a \mid c)$

a. True. By the product rule we know $P(b, c)P(a|b, c) = P(a, c)P(b|a, c)$, which by assumption reduces to $P(b, c) = P(a, c)$. Dividing through by $P(c)$ gives the result.
In other words,
$P(a, b, c) = P(b, c)P(a|b, c) = P(a, c)P(b|a, c)$.
Using $P(a|b, c) = P(b|a, c)$, we get $P(b, c)P(b|a, c) = P(a, c)P(b|a, c)$, thus $P(b, c) = P(a, c)$.
Thus $P(b, c)/P(c) = P(a, c)/P(c)$, so $P(b|c) = P(a|c)$.

b. False. The statement $P(a|b, c) = P(a)$ merely states that a is independent of b and c, it makes no claim regarding the dependence of b and c. A counter-example: a and b record the results of two independent coin flips, and c = b.

c. False. While the statement $P(a|b) = P(a)$ implies that a is independent of b, it does not imply that a is conditionally independent of b given c. A counter-example: a and b record the results of two independent coin flips, and c equals the xor of a and b.

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Figure 13.3 A full joint distribution for the *Toothache*, *Cavity*, *Catch* world.

13.8 Given the full joint distribution shown in Figure 13.3, calculate the following:

a. $\mathbf{P}(toothache)$.

b. $\mathbf{P}(Cavity)$.

c. $\mathbf{P}(Toothache \mid cavity)$.

d. $\mathbf{P}(Cavity \mid toothache \lor catch)$.

The main point of this exercise is to understand the various notations of uppercase versus lowercase variable names.
The rest is easy, involving a small matter of addition.

a. This asks for the probability that Toothache is true.
P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

b. This asks for the vector of probability values for the random variable Cavity. It has two values, which we list in the order <true, false>.
First add up 0.108 + 0.012 + 0.072 + 0.008 = 0.2. Then we have P(Cavity) = <0.2, 0.8>.

c. This asks for the vector of probability values for Toothache, given that cavity is true.
P(Toothache|cavity) = <(.108 + .012)/0.2, (0.072 + 0.008)/0.2> = <0.6, 0.4>

d.This asks for the vector of probability values for Cavity, given that either Toothache or Catch is true.
First compute P(toothache ∨ catch) = 0.108+0.012+0.016+0.064+ 0.072 + 0.144 = 0.416. (the sum of the first 3 columns).
Then P(Cavity| toothache ∨ catch) = <(0.108 + 0.012 + 0.072)/0.416, (0.016 + 0.064 + 0.144)/0.416> = <0.4615, 0.5384>

13.14 Suppose you are given a coin that lands *heads* with probability x and *tails* with probability $1 - x$. Are the outcomes of successive flips of the coin independent of each other given that you know the value of x? Are the outcomes of successive flips of the coin independent of each other if you do *not* know the value of x? Justify your answer.

If the probability x is known, then successive flips of the coin are independent of each other, since we know that each flip of the coin will land heads with probability x. Formally, if F1 and F2 represent the results of two successive flips, we have

$P(F1 = \text{heads}, F2 = \text{heads}|x) = x * x = P(F1 = \text{heads}|x)P(F2 = \text{heads}|x)$

Thus, the events F1 = heads and F2 = heads are independent.

If we do not know the value of x, however, the probability of each successive flip is dependent on the result of all previous flips. The reason for this is that each successive flip gives us information to better estimate the probability x (i.e., determining the posterior estimate for x given our prior probability and the evidence we see in the most recent coin flip). This new estimate of x would then be used as our "best guess" of the probability of the coin coming up heads on the next flip. Since this estimate for x is based on all the previous flips we have seen, the probability of the next flip coming up heads depends on how many heads we saw in all previous flips, making them dependent.

For example, if we had a uniform prior over the probability x, then one can show that after n flips if m of them come up heads then the probability that the next one comes up heads
is $(m + 1)/(n + 2)$, showing dependence on previous flips.

13.13 Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

Let V be the statement that the patient has the virus, and A and B the statements that the medical tests A and B returned positive, respectively. The problem statement gives:

$P(V) = \langle 0.01, 0.99 \rangle$

$P(A|v) = \langle 0.95, 0.05 \rangle$

$P(A|\neg v) = \langle 0.10, 0.90 \rangle$

$P(B|v) = \langle 0.90, 0.10 \rangle$

$P(B|\neg v) = \langle 0.05, 0.95 \rangle$

The test whose positive result is more indicative of the virus being present is the one whose posterior probability, $P(V \mid a)$ or $P(V \mid b)$ is largest.
One can compute these probabilities directly from the information given, finding that $P(v \mid a) = 0.0876$ and $P(v \mid b) = 0.1538$, so B is more indicative.

$P(V \mid a) = P(a \mid V)P(V) / P(a) = \langle P(a \mid v)P(v), P(a \mid \neg v)P(\neg v) \rangle / P(a) = \langle 0.95*0.01, 0.10*0.99 \rangle / P(a) = \langle 0.0095, 0.099 \rangle / P(a) = \langle 0.0876, 0.9124 \rangle$

$P(V \mid b) = P(b \mid V)P(V) / P(b) = \langle P(b \mid v)P(v), P(b \mid \neg v)P(\neg v) \rangle / P(b) = \langle 0.90*0.01, 0.05*0.99 \rangle / P(a) = \langle 0.009, 0.0495 \rangle / P(a) = \langle 0.1538, 0.8462 \rangle$

13.17 Show that the statement of conditional independence

$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z)$$

is equivalent to each of the statements

$$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z) \quad \text{and} \quad \mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z)\,.$$

$$\begin{aligned} P(X \mid Y, Z) &= P(X, Y, Z) / P(Y, Z) \\ &= P(X, Y \mid Z)\, P(Z) / P(Y, Z) \\ &= P(X \mid Z)\, P(Y \mid Z)\, P(Z) / P(Y, Z) \\ &= P(X \mid Z)\, P(Y, Z) / P(Y, Z) \\ &= P(X \mid Z) \end{aligned}$$

Similarly, $P(Y \mid X, Z) = P(Y \mid Z)$