Worksheet Week 20

Problems

Q1. X and Y are two (binary) random variables. If X and Y are independent, then $P(X,Y) = P(X)P(Y)$

(a) Give an example of two random variables that are independent.

For example - two flips of a coin. Suppose we have X = coin 1 and Y = coin 2. Then $P(X = head, Y = head) = P(X = head)P(Y = head)$

(b) Complete the probability table below in such way that the variables X and Y are independent.

	$X = 0$	$X = 1$
$Y = 0$		
$Y = 1$		

Solution:
For example,

	$X = 0$	$X = 1$
$Y = 0$	0.2	0.3
$Y = 1$	0.2	0.3

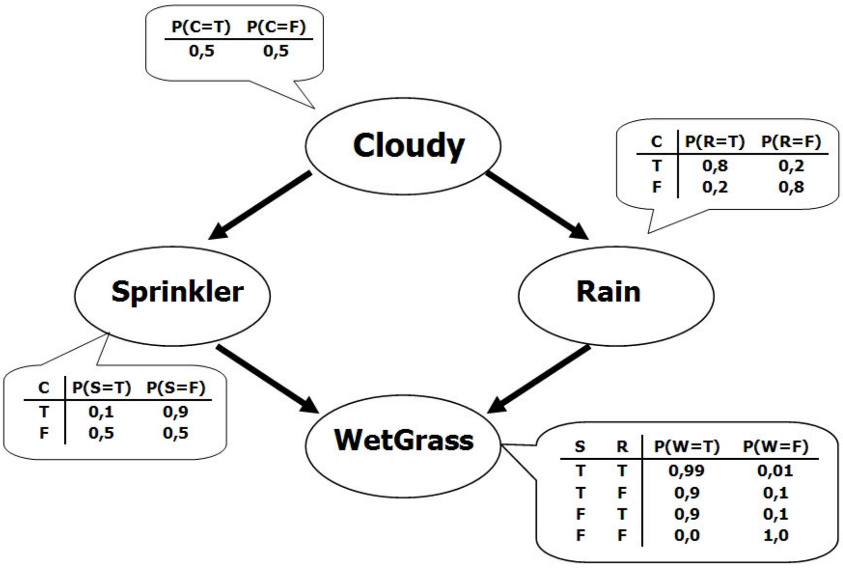
Marginalising over X, we get $P(X = 0) = 0.4$, $P(X = 1) = 0.6$. Marginalising over Y we get $P(Y = 0) = P(Y = 1) = 0.5$. We see that $P(X,Y) = P(X)P(Y)$ for each combination of X and Y.

(c) Determine the missing entries (a, b) of the joint distribution in such a way that the variables X and Y are again independent.

$$P(Y = 0, X = 0) = 0.1$$
$$P(Y = 0, X = 1) = 0.3$$
$$P(Y = 1, X = 0) = a$$
$$P(Y = 1, X = 1) = b$$

Solution: We are given

	$X = 0$	$X = 1$
$Y = 0$	0.1	0.3
$Y = 1$	a	b



$P(Y=0) = 0.4$ so $P(Y=1) = 1-0.4 = 0.6$. We need $P(X=0, Y=1) = P(X=0)P(Y=1)$, $P(X=1, Y=1) = P(X=1)P(Y=1)$. So, $0.4*(0.1+a) = 0.1$ and $0.4*(0.3+b) = 0.3$, Therefore $0.1+a = 0.1/0.4 = 0.25$ and hence $a = 0.15$. $0.3+b = 0.3/0.4 = 0.75$ and hence $b = 0.45$

Q2. Solve the questions on slides 42 and 44 of the lecture slides. What is the marginal probability, P(S=1), that the sprinkler is on, given that the grass is wet?

$$P(S=1|W=1) = \frac{P(S=1, W=1)}{P(W=1)}$$

$$P(W=1) = \sum_R \sum_C \sum_S P(W=1, S, C, R)$$

$$P(S=1, W=1) = \sum_R \sum_C P(W=1, S=1, C, R)$$

The joint probability $P(W, S, C, R)$ is given by

$$P(W, S, C, R) = P(C)P(S|C)P(R|C)P(W|S, R) \quad (1)$$

So,

$$\begin{aligned}
P(W=1) &= \sum_C \sum_S \sum_R P(C)P(S|C)P(R|C)P(W=1|S, R) && (2)\\
&= 0.5*0.1*0.8*0.99 + 0.5*0.1*0.2*0.9 && (3)\\
&\quad + 0.5*0.9*0.8*0.9 + 0.5*0.9*0.2*0 && (4)\\
&\quad + 0.5*0.5*0.2*0.99 + 0.5*0.5*0.8*0.9 && (5)\\
&\quad + 0.5*0.5*0.2*0.9 + 0.5*0.5*0.8*0 && (6)\\
&= 0.6471 && (7)
\end{aligned}$$

$$P(W=1,S=1) = \sum_R \sum_C P(C)P(S=1|C)P(R|C)P(W=1|S=1,R) \quad (8)$$
$$= 0.5*0.1*0.8*0.99 + 0.5*0.1*0.2*0.9 \quad (9)$$
$$+ 0.5*0.5*0.2*0.99 + 0.5*0.5*0.8*0.9 \quad (10)$$
$$= 0.2781 \quad (11)$$

So $P(S=1|W=1) = \frac{P(S=1,W=1)}{P(W=1)} = 0.2781/0.6471 = 0.4297$

What is the posterior probability, $P(S=1|W=1,R=1)$, that the sprinkler is on given that the grass is wet and it is raining?

$$P(S=1|W=1,R=1) = \frac{P(S=1,W=1,R=1)}{P(W=1,R=1)}$$

$$P(W=1,R=1) = \sum_S \sum_C P(S,W=1,R=1,C)$$

$$P(S=1,W=1,R=1) = \sum_C P(S=1,W=1,R=1,C)$$

$$P(W=1,R=1) = \sum_C \sum_S P(C)P(S|C)P(R=1|C)P(W=1|S,R=1) \quad (12)$$
$$= 0.5*0.1*0.8*0.99 + 0.5*0.9*0.8*0.9 \quad (13)$$
$$+ 0.5*0.5*0.2*0.99 + 0.5*0.5*0.2*0.9 \quad (14)$$
$$= 0.4581 \quad (15)$$

$$P(W=1,R=1,S=1) = \sum_C P(C)P(S=1|C)P(R=1|C)P(W=1|S=1,R=1) \quad (16)$$
$$= 0.5*0.1*0.8*0.99 + 0.5*0.5*0.2*0.99 \quad (17)$$
$$= 0.0846 \quad (18)$$

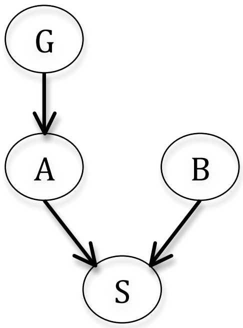
So $P(S=1|W=1,R=1) = 0.0846/0.4581 = 0.1846$

I.e., if we know it is raining, this decreases the probability that the sprinkler is on, given that the ground is wet.

Q3. A patient can have a symptom, S, that is caused by two different diseases, A and B. It is known that the presence of a gene G is important in the manifestation of disease A. The Bayes net and conditional probability tables are shown in Figure 2.

(a) What is the probability that a patient has disease A?

$$P(a) = P(a|g)P(g) + P(a|\neg g)P(\neg g) \quad (19)$$
$$= 1*0.1 + 0.1*0.9 = 0.19 \quad (20)$$



$P(G)$	
g	0.1
$\neg g$	0.9

$P(B)$	
b	0.4
$\neg b$	0.6

$P(A\|G)$		
g	a	1.0
g	$\neg a$	0.0
$\neg g$	a	0.1
$\neg g$	$\neg a$	0.9

$P(S\|A,B)$			
a	b	s	1.0
a	b	$\neg s$	0.0
a	$\neg b$	s	0.9
a	$\neg b$	$\neg s$	0.1
$\neg a$	b	s	0.8
$\neg a$	b	$\neg s$	0.2
$\neg a$	$\neg b$	s	0.1
$\neg a$	$\neg b$	$\neg s$	0.9

Figure 1: Bayes net and probability tables for Q5

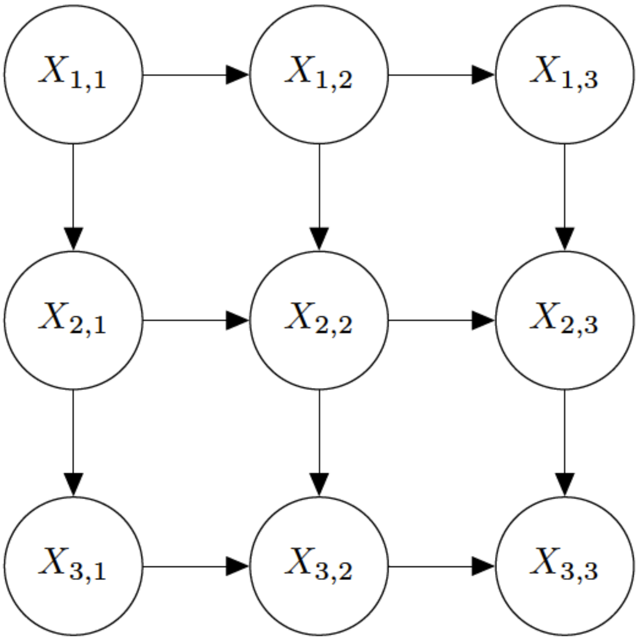
(b) What is the probability that a patient has disease A if we know that the patient has disease B

A and B are conditionally independent, so $P(a|b) = P(a) = 0.19$

(c) What is the probability that a patient has disease A if we know that the patient has disease B AND symptom S

$$
\begin{aligned}
P(a|s,b) &= \frac{P(a,s,b)}{P(s,b)} && (21) \\
&= \frac{P(a)P(b)P(s|a,b)}{P(a)P(b)P(s|a,b) + P(\neg a)P(b)P(s|\neg a,b)} && (22) \\
&= \frac{0.19 * 0.4 * 1}{0.19 * 0.4 * 1 + 0.81 * 0.4 * 0.8} = 0.2267 && (23)
\end{aligned}
$$

Q4. Consider the following Bayesian network:



(a) Which random variables are independent of $X_{3,1}$?

Solution:

None. Recall that two random variables X and Y are independent if $P(X, Y) = P(X)P(Y)$. In a diagram, this equates to having two disconnected nodes. In the diagram above, consider writing out the joint distribution and then marginalising to obtain e.g. $P(X_{3,1}, X_{2,2})$. As we marginalise, we can eliminate some variables, e.g. $X_{3,3}, X_{2,3}, X_{1,3}, X_{3,2}$, but we will not be able to end up with simply $P(X_{3,1}, X_{2,2}) = P(X_{3,1})P(X_{2,2})$

A comprehensive algorithm for determining independence is called d-separation, and you can see this at `http://web.mit.edu/jmn/www/6.034/d-separation.pdf`

(b) Which random variables are independent of $X_{3,1}$ given $X_{1,1}$?

Solution:

Both $X_{1,2}$ and $X_{1,3}$ become independent of $X_{3,1}$ given $X_{1,1}$.