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

   (a) If $P(a|b,c) = P(b|a,c)$, then $P(a|c) = P(b|c)$
   (b) If $P(a|b,c) = P(a)$, then $P(b|c) = P(b)$
   (c) If $P(a|b) = P(a)$, then $P(a|b,c) = P(a|c)$

Identities

Probability of a *given* b

$$\rightarrow \quad P(a|b) = \frac{P(a,b)}{P(b)}$$

Probability of a *and* b

$$P(a,b) = P(a|b)P(b)$$

Bayes formula (flip the order of given variable)

$$P(b|a) = P(a|b) \cdot \frac{P(b)}{P(a)}$$

1)

$$P(a|b,c) = P(b|a,c)$$

$$\Rightarrow \frac{P(a,b,c)}{P(b,c)} = \frac{P(a,b,c)}{P(a,c)}$$

$$\Rightarrow P(b,c) = P(a,c)$$

$$\Rightarrow P(b|c)\cdot \cancel{P(c)} = P(a|c)\cdot \cancel{P(c)}$$

$$\Rightarrow P(b|c) = P(a|c) \quad \checkmark$$

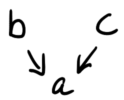
2)

$$P(a|b,c) = P(a)$$

$\Rightarrow$ b and c are independent from a.
But there is no information about b being independent from c

$$\not\Rightarrow P(b|c) = P(c) \quad \times$$

We can also show this by drawing the corresponding bayesian network:

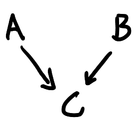


Here it is evident that b and c are within eachothers Markov blanket, meaning that they depend on each other. Thereby disprobing the independence of b and c.

c)

If c is an event that depends on a and b. a and b will not be independent. Therfore, this statement is not always true. $\times$

This can also be seen by drawing the bayesian network where c is dependent on a and b.



2. Given the full joint distribution shown below (Figure 13.3), calculate the following

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

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

   (c) $\mathbf{P}(Cavity|catch)$.

   (d) $\mathbf{P}(Cavity|toothache \vee catch)$.

.

	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.

a) $P(toothache) = 0{,}108 + 0{,}012 + 0{,}016 + 0{,}064 = 0{,}20$

b) $P(Catch) = P(catch, \neg catch)$

$$\begin{cases} P(catch) = 0{,}108 + 0{,}016 + 0{,}072 + 0{,}144 = 0{,}34 \\ P(\neg catch) = 0.012 + 0.064 + 0.008 + 0.576 = 0{,}66 \end{cases}$$

c) $P(Cavity|catch) = P(cavity|catch, \neg cavity|catch)$

$$= \begin{cases} P(cavity|catch) = 0.108 + 0.072 = 0.18 \\ P(\neg cavity|catch) = 0.016 + 0.144 = 0.16 \end{cases}$$

d) $P(Cavity|toothache, catch) = P(cavity|\{toothache, catch\}, \neg cavity|\{toothache, catch\})$

$$= \begin{cases} P(cavity|toothache, catch) = 0.108 \\ P(\neg cavity|toothache, catch) = 0.016 \end{cases}$$

3. We wish to transmit an n-bit message to a receiving agent. The bits in the message are independently corrupted (flipped) during transmission with ϵ probability each. With an extra parity bit sent along with the original information, a message can be ~~corrected~~ Discarded by the receiver if at most one bit in the entire message (including the parity bit) has been corrupted. Suppose we want to ensure that the correct message is received with probability at least $1 - \delta$. What is the maximum feasible value of n? Calculate this value for the case $\epsilon = 0.002$, $\delta = 0.01$.

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} \theta(x_i | parents(X_i)) \tag{14.1}$$

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \tag{13.3}$$

$$\mathbf{P}(Y) = \sum_{z \in Z} P(Y, z) \tag{13.6}$$

Parity

There will be n+1 random variables in the network.

Chance for each variable to be true: $1 - \epsilon = 0{,}998$

$$\overset{0{,}998}{X_1} \rightarrow \overset{0{,}998}{X_2} \rightarrow \ldots \rightarrow \overset{0{,}998}{X_{n+1}}$$

We can then calculate the network output probability

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} \underbrace{P(x_i | Parents(X_i))} = \prod_{i=1}^{n+1} (1-\epsilon)^i = (1-\epsilon)^{n+1}$$

Propability of each node being true independent of its parents.
In out case this is the same for all nodes.

$$P(C=0) = (1-\epsilon)^{n+1}$$

Propability of ONE bit being corrupted

$$P(C=1) = (1-\epsilon)^n \cdot \overset{\text{corrupted}}{\epsilon} \cdot \underbrace{(n+1)}_{\text{Number of bits}}$$

Propability of either

$$P(C=0, C=1) = P(C=0) + P(C=1) = (1-\epsilon)^{n+1} + (1-\epsilon)^n \cdot \epsilon \cdot (n+1)$$

Solving for n

$$P(C=0, C=1) = 1 - \delta$$

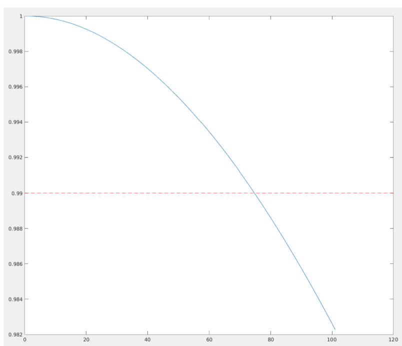
```
syms n

eta = 0.002;
sigma = 0.01;

f = (1-eta)^(n+1) + (1-eta)^n * eta * (n+1);

assume(n > 0);
sol = solve(f == 1-sigma, n);
floor(sol) % => 73
```

$\Rightarrow \underline{\underline{n = 73}}$



4. Equation (13.1) on page 433 defines the joint distribution represented by a Bayesian network in terms of the parameter $\theta(X_i|Parents(X_i))$. This exercise asks you to derive the equivalence between the parameters and the conditional probabilities $\mathbf{P}(X_i|Parents(X_i))$ from this definition. (For part **c**, show the resulting expression reduces to $\theta(z|y)$ to be consistent with other parts of the problem.)

(a) Consider a simple network $X \to Y \to Z$ with three Boolean variables. Use Equations (13.3) and (13.6) (pages 434 and 455) to express the conditional probability $P(z|y)$ as the ratio of two sums, each over entries in the joint distribution $\mathbf{P}(X, Y, Z)$.

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} \theta(x_i|parents(X_i)) \tag{14.1}$$

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \tag{13.3}$$

$$\mathbf{P}(Y) = \sum_{z \in Z} P(Y, z) \tag{13.6}$$

$$X \to Y \to Z$$

$$P(z|y) = \frac{P(z,y)}{P(y)} = \frac{\sum_x P(x,y,z)}{\sum_x \sum_y P(x,y,z)}$$

(b) Now use Equation (13.1) to write this expression in terms of the network parameters $\theta(X)$, $\theta(Y|X)$, and $\theta(Z|Y)$.

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} \theta(x_i|parents(X_i))$$

$$P(z|y) = \frac{\sum_x \theta(x)\theta(y|x)\theta(z|y)}{\sum_x \sum_y \theta(x)\theta(y|x)\theta(z|y)}$$

(c) Next, expand out the summations in your expression from part (b), writing out explicitly the terms for the true and false values of each summed variable. Assuming that all network parameters satisfy the constraint $\sum_{x_i} \theta(x_i|Parents(X_i)) = 1$, show that the resulting expression reduces to $\theta(z|y)$.

I don't get this...

$$P(z|y) = \frac{\sum_x \theta(x)\theta(y|x)\theta(z|y)}{\sum_x \sum_y \theta(x)\theta(y|x)\theta(z|y)} = \frac{\theta(x_t)\theta(y|x_t)\theta(z|y) + \theta(x_f)\theta(y|x_f)\theta(z|y)}{\sum_y \theta(x_t)\theta(y|x_t)\theta(z|y) + \theta(x_f)\theta(y|x_f)\theta(z|y)}$$

$$\Rightarrow P(z|y) = \frac{\theta(x_t)\theta(y|x_t)\theta(z|y) + \theta(x_f)\theta(y|x_f)\theta(z|y)}{\theta(x_t)\theta(y_t|x_t)\theta(z|y_t) + \theta(x_f)\theta(y_t|x_f)\theta(z|y_t) + \theta(x_t)\theta(y_f|x_t)\theta(z|y_f) + \theta(x_f)\theta(y_f|x_f)\theta(z|y_f)}$$

(d) Generalize this derivation to show that $\theta(X_i|Parents(X_i)) = \mathbf{P}(X_i|Parents(X_i))$ for any Bayesian network.

5. Consider the Bayesian network in Figure 14.2.

(a) If no evidence is observed, are *Burglary* and *Earthquake* independent? Prove this from the numerical semantics and from the topological semantics.

Both "Burglary" and "Earthquake" nodes have no parent nodes, and must therefore be independent from any other node.

This can also be observed as their probability table does not include any outside outcomes.

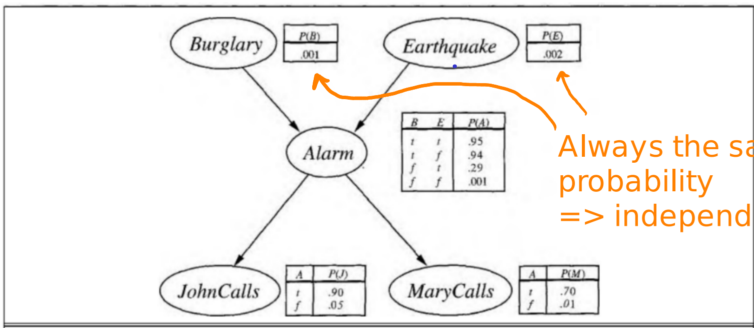


Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

Always the same probability => independent

(b) If we observe *Alarm=true*, are *Burglary* and *Earthquake* independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence.

They should be dependent as they are within each others markov blanket

$$\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X,\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X,\mathbf{e},\mathbf{y})$$

$\mathbf{E}$: Evidence variables

$\mathbf{Y}$: Unobserved Variables

X: Query variable

α: $1/P(\mathbf{E})$

We are trying to prove the following statement:

$$P(B|A,E) \neq P(B|A)$$

Definition of independence

$$P(B|A,E) = \alpha P(B,A,E) = \alpha \sum_J \sum_M P(B,A,E,J,M)$$

$$\Rightarrow P(B|A,E) = \alpha \sum_J \sum_M P(B)P(E)P(A|B,E)P(J|A)P(M|A)$$

$$\Rightarrow P(B|A,E) = \alpha P(B)P(E)P(A|B,E)\cancel{\sum_J} P(J|A)\cancel{\sum_M} P(M|A)$$

$$\Rightarrow P(B|A,E) = \alpha \cdot 0{,}01 \cdot 0{,}02 \cdot 0{,}95 \cdot 0{,}90 \cdot 0{,}70$$

$$P(a,b) = P(a|b)P(b)$$

$$\alpha = \frac{1}{P(A,E)} = \frac{1}{P(A|E)P(E)} = \frac{1}{(0{,}95 \cdot 0{,}29)\cdot 0{,}02} \approx 181{,}5$$

Is this correct??

$$\Rightarrow P(B|A,E) \approx 0{,}02172$$

Alright, now lets try without earthquake as a given!

$$P(\overset{X}{B}|\overset{E}{A}) = \alpha P(B,A) = \alpha \cdot \sum_A \sum_J \sum_M \overbrace{P(B)P(E)P(A|B,E)P(J|A)P(M|A)}^{\text{Network}}$$

$$P(B|A) = \alpha P(B)P(E)P(A|B,E)\sum_A \cancel{\sum_J \sum_M} P(J|A)P(M|A)$$

$$P(B|A) = \alpha \cdot P(B)P(E)P(A|B,E)\sum_{A} P(J|A)P(M|A)$$

$$P(B|A) = \alpha \cdot 0{,}01 \cdot 0{,}02 \cdot 0{,}95 \cdot (0{,}90 \cdot 0{,}70 + 0{,}05 \cdot 0{,}01)$$

$$\alpha = \frac{1}{P(A)} = \frac{1}{0{,}95 + 0{,}94 + 0{,}29 + 0{,}01} \approx 386{,}1$$

$$P(B|A) \approx 0{,}4625$$

Comparing the two results

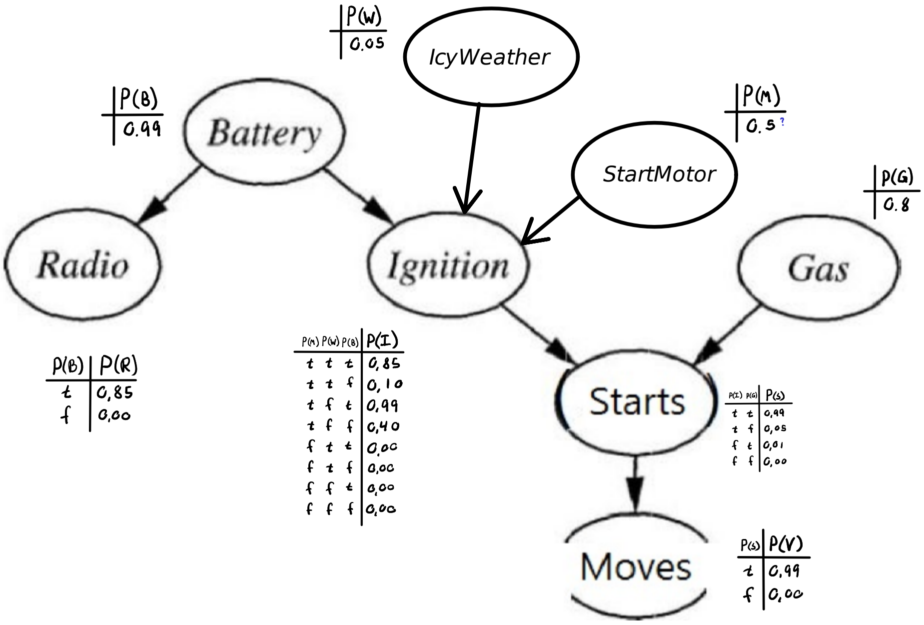
$$\begin{cases} P(B|A,E) \approx 0{,}02172 \\ P(B|A) \approx 0{,}4625 \end{cases} \Rightarrow P(B|A,E) \neq P(B|A) \Rightarrow \text{DEPENDENT!}$$

6. Consider the network for car diagnosis shown in Figure 14.21.

(a) Extend the network with the Boolean variables *IcyWeather* and $\underbrace{\textit{StarterMotor}}$.

I dont know what this is supposed to mean...

(b) Give reasonable conditional probability tables for all the nodes.



P(W)
0.05

P(B)
0.99

P(M)
0.5 ?

P(G)
0.8

P(B)	P(R)
t	0,85
f	0,00

P(M)	P(W)	P(B)	P(I)
t	t	t	0,85
t	t	f	0,10
t	f	t	0,99
t	f	f	0,40
f	t	t	0,00
f	t	f	0,00
f	f	t	0,00
f	f	f	0,00

P(I)	P(G)	P(S)
t	t	0,99
t	f	0,05
f	t	0,01
f	f	0,00

P(S)	P(V)
t	0,99
f	0,00

(c) How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming that no conditional independence relations are known to hold among them?

(d) How many independent probability values do your network tables contain?

The network contains 4 independent variables: B, W, M and G

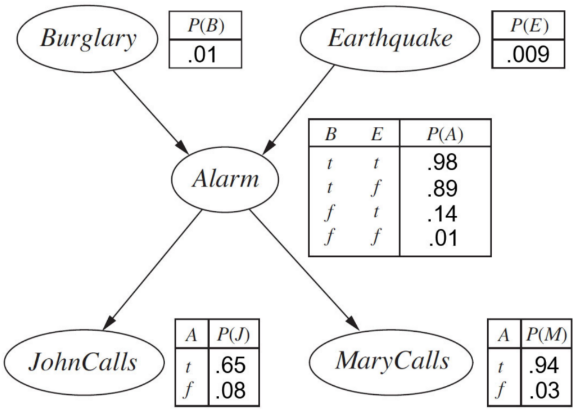
(e) The conditional distribution for *Starts* could be described as a **noisy-AND** distribution. Define this family in general and relate it to the noisy-OR distribution.

In a noisy-AND distribution, ALL the inputs must be true for the output to have a posibility of being true.

In a noisy-OR distrybution, SOME inputs must be true for hte output to have a posibility of being true.

7. Consider the modified burglary network shown in Figure 1.

(a) Implement the variable elimination algorithm for computing $P(B|j,m)$ using MATLAB. (Turn in a printout of your MATLAB code.)



P(B)
.01

P(E)
.009

B	E	P(A)
t	t	.98
t	f	.89
f	t	.14
f	f	.01

A	P(J)
t	.65
f	.08

A	P(M)
t	.94
f	.03

Figure 1: Modified burglary network.

$$P(B|j,m) = \alpha \cdot \sum_{a,e} P(B,j,m,a,e)$$

$$= \alpha \cdot \sum_{a,e} P(B) P(e) P(a|B,e) P(j|a) P(m|a)$$

$$= \alpha \cdot \underbrace{P(B)}_{f_1} \cdot \sum_{e} \underbrace{P(e)}_{f_2} \sum_{a} \underbrace{P(a|B,e)}_{f_3} \underbrace{P(j|a)}_{f_4} \underbrace{P(m|a)}_{f_5}$$

$$= \alpha \cdot \underbrace{P(B)}_{f_1} \cdot \sum_e \underbrace{P(e)}_{f_2} \sum_a \underbrace{P(a|B,e)}_{f_3} \underbrace{P(j|a)}_{f_4} \underbrace{P(m|a)}_{f_5}$$

B	$f_1(B)$
t	0.01
f	0.99

E	$f_2(E)$
t	0.009
f	0.991

A	B	E	$f_3(A,B,E)$
t	t	t	0.98
t	t	f	0.89
t	f	t	0.14
t	f	f	0.01
f	t	t	0.02
f	t	f	0.11
f	f	t	0.86
f	f	f	0.99

J	A	$f_4(J,A)$
t	t	0.65
t	f	0.08
f	t	0.35
f	f	0.92

M	A	$f_5(M,A)$
t	t	0.94
t	f	0.03
f	t	0.06
f	f	0.97

A	f_6
t	0.65
f	0.08

A	f_7
t	0.94
f	0.03

$$\Rightarrow P(B|j,m) = \alpha \cdot f_1(B) \sum_e f_2(e) \sum_a f_3(a,B,e)\, f_4(j,a)\, f_5(j,a)$$

Restrict

$$\Rightarrow P(B|j,m) = \alpha \cdot f_1(B) \sum_e f_2(e) \sum_a \overbrace{f_3(a,B,e)\, f_6(a)\, f_7(a)}^{f_8}$$

Multiply

$$= \alpha \cdot f_1(B) \sum_e f_2(e) \sum_a f_8(a,B,e)$$

Sum-Out

$$= \alpha \cdot f_1(B) \sum_e f_2(e) \overbrace{\left(f_8(a,B,e) + f_8(\neg a,B,e)\right)}^{f_9}$$

$$= \alpha \cdot f_1(B) \sum_e \overbrace{f_2(e)\, f_9(B,e)}^{f_{10}}$$

Multiply

$$= \alpha \cdot f_1(B) \sum_e f_{10}(B,e)$$

Sum-Out

$$= \alpha \cdot f_1(B) \cdot \overbrace{\left(f_{10}(B,e) + f_{10}(B,\neg e)\right)}^{f_{11}}$$

$$= \alpha \cdot \overbrace{f_1(B) \cdot f_{11}(B)}^{f_{12}}$$

Multiply

$$= \alpha \cdot f_{12}(B)$$

A	B	E	$f_8 = f_3 f_6 f_7$
t	t	t	0.98 · 0.65 · 0.94 = 0.600
t	t	f	0.89 · 0.65 · 0.94 = 0.544
t	f	t	0.14 · 0.65 · 0.94 = 0.086
t	f	f	0.01 · 0.65 · 0.94 = 0.006
f	t	t	0.02 · 0.08 · 0.03 = 0.000
f	t	f	0.11 · 0.08 · 0.03 = 0.000
f	f	t	0.86 · 0.08 · 0.03 = 0.002
f	f	f	0.99 · 0.08 · 0.03 = 0.002

B	E	f_9
t	t	0.600 + 0.000 = 0.600
t	f	0.544 + 0.000 = 0.544
f	t	0.086 + 0.002 = 0.088
f	f	0.006 + 0.002 = 0.008

B	E	$f_{10} = f_2 \cdot f_9$
t	t	0.600 · 0.009 = 0.005
t	f	0.544 · 0.991 = 0.539
f	t	0.088 · 0.009 = 0.001
f	f	0.008 · 0.991 = 0.008

B	f_{11}
t	0.005 + 0.539 = 0.544
f	0.001 + 0.008 = 0.009

B	$f_{12} = f_{11} \cdot f_1$
t	0.544 · 0.01 = 0.005
f	0.009 · 0.99 = 0.009

$$\Rightarrow P(B|j,m) = \text{norm}\left(\begin{bmatrix} 0.005 \\ 0.009 \end{bmatrix}\right) = \begin{bmatrix} 0.36 \\ 0.64 \end{bmatrix}$$

(b) Find $P(B|J,M)$ and $P(E|J,M)$ using the program implemented in (a). There are a total of 16 probabilities. (Turn in a printout of the outputs from your MATLAB code. You may use the command `diary` in MATLAB.)

See attached file 2024-81380-matlab.zip

Matlab Output

```
%% General Implementation
P(B | J = true, M = true)

B
1.000 0.374
0.000 0.626

P(B | J, M)

BJM
1.000 1.000 1.000 0.005
1.000 1.000 0.000 0.000
1.000 0.000 1.000 0.003
1.000 0.000 0.000 0.001
0.000 1.000 1.000 0.009
0.000 1.000 0.000 0.076
0.000 0.000 1.000 0.031
0.000 0.000 0.000 0.874

P(E | J, M)

EJM
1.000 1.000 1.000 0.001
1.000 1.000 0.000 0.001
1.000 0.000 1.000 0.001
1.000 0.000 0.000 0.007
0.000 1.000 1.000 0.014
0.000 1.000 0.000 0.076
0.000 0.000 1.000 0.033
0.000 0.000 0.000 0.868
```

8. Consider the modified weather network shown in Figure 2.

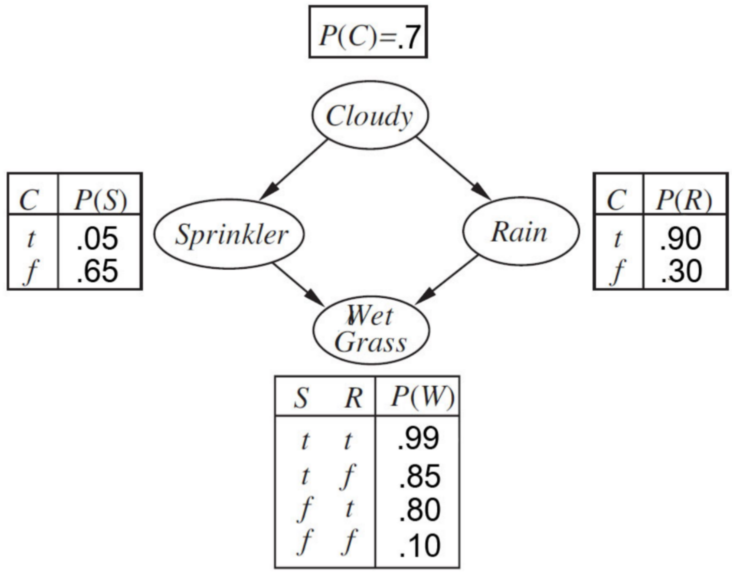


Figure 2: Modified weather network.

(a) Is S and R independent? Is S and R conditionally independent given C? Prove your answers from the numerical semantics.

$$P(S|c) = \alpha \cdot \sum_{W,R} P(S,c,W,R) = \alpha \cdot \sum_{W,R} P(c)P(S|c)P(R|c)P(W|S,R)$$

$$= \alpha \cdot P(c)P(S|c) \sum_R P(R|c) \sum_W P(W|S,R)$$

$$= \alpha \cdot \overset{0.7}{P(c)}\overset{\begin{bmatrix}0.05\\0.95\end{bmatrix}}{P(S|c)} \sum_R P(R|c)\left(P(w|S,R) + P(\neg w|S,R)\right)$$

$$= \alpha \cdot \begin{bmatrix}0.35\\0.65\end{bmatrix} \cdot \left[\left(\cancel{\overset{0.90}{P(r|c)}}\left(\overset{0.99}{P(w|s,r)} + \overset{0.01}{P(\neg w|s,r)}\right)\right) + \cancel{\overset{0.10}{P(\neg r|c)}}\left(\overset{0.85}{P(w|s,\neg r)} + \overset{0.15}{P(\neg w|s,\neg r)}\right)\right]$$

$$\Rightarrow P(S,c) = \begin{bmatrix}0.05\\0.95\end{bmatrix}$$

$$P(S|c,r) = \alpha \cdot P(c)P(S|c)\,P(r|c) \sum_W P(W|S,r)$$

$$= \alpha \cdot \overset{0.7}{P(c)}\overset{\begin{bmatrix}0.05\\0.95\end{bmatrix}}{P(S|c)}\,\overset{0.90}{P(r|c)}\left(\overset{0.99}{P(w|S,r)} + \overset{0.01}{P(\neg w|S,r)}\right)$$

$$\Rightarrow P(S|c,r) = \begin{bmatrix}0.05\\0.95\end{bmatrix}$$

$\Rightarrow P(S|c) = P(S|c,r) \Rightarrow$ **P(S|C) is INDPENDENT from P(S|C,R)**

(c) Implement the direct sampling algorithm to estimate $P(S, R|W)$ using MATLAB. You need to use an enough number of samples for good estimates for the exact values. How many samples does it need to achieve an accuracy of ± 0.01? (Turn in printouts of your MATLAB code and outputs.)

I gave it a try, but did not succeed. See file 'problem8.m'.