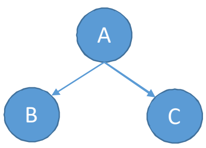
Exercise 5

1. Compare and explain the different sampling methods by giving an example:
   a. Prior sampling
   b. Rejection sampling
   c. Weighted sampling

   and their differences.

Prior Sampling:
sampling an empty network. Generate examples (topological order) based on the known local distribution (CPTs). Used to compute prior probabilities, e.g., P(C)



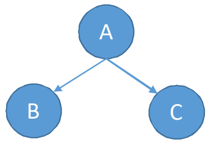
For i=1 to N

- Sample a value a from CPT_A
- Sample a value b from CPT_B given a
- Sample a value c from CPT_C given a

1. Compare and explain the different sampling methods by giving an example:
   a. Prior sampling
   b. Rejection sampling
   c. Weighted sampling

   and their differences.

Rejection Sampling:
Create samples with Prior sampling. Throw away samples that do not agree with the evidences.



For i=1 to N

- Sample a value a from CPT_A
- Sample a value b from CPT_B given a
- Sample a value c from CPT_C given a

Used to compute probabilities given some evidences , e.g., P(C|A=true)
Throw away samples where A=false.

3

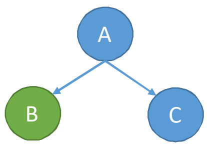
1. Compare and explain the different sampling methods by giving an example:
   a. Prior sampling
   b. Rejection sampling
   c. Weighted sampling

   and their differences.

Weighted sampling :
Include knowledge of known evidences in the generation process. Use this knowledge for generating the values of the childs. Unknown variables are randomly generated by including knowledge of the parents. Known variables are fixed but a weight is computed that reflects how well the evidence value of the known variable fit to the rest of the values generated so far. Only includes influence of parents.

Suppose B=true, an observation

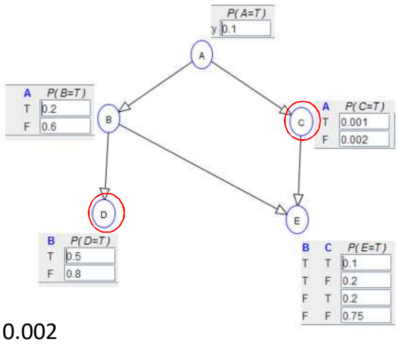
For i=1 to N



- Sample a value a from CPT_A, suppose A=false
- Weight the example by: w = w*P(B=true| A=false)
- Sample a value c from CPT_C given a=false

4

2. Given the following network. Generate a possible example for P(A|C,¬D). Also compute all possible weights for examples generated with this observations. Are there any irrelevant attributes that would be pruned before evaluating the query with variable elimination?



The possible weights are generated by the variables of the evidences, C and ¬D.

The weight formula is $w = w_c * w_{\neg D}$

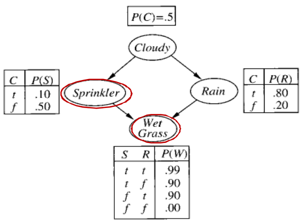
For w_c: if A=true the weight is 0.001. If A is false the weight is 0.002

For $w_{\neg D}$: If B=true the weight is 0.5. If B is false the weight is 0.2

There are four combinations for w given the possible values for parents of C and ¬D.

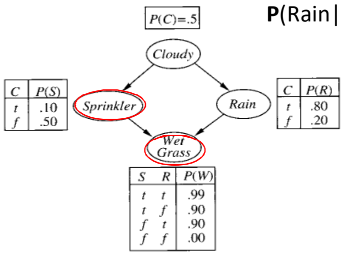
There is one variable that can be pruned: E.

3. Consider the query **P**(Rain| Sprinkler=true, WetGrass=true) (see figure below) and how Gibbs can answer it.



a. How many states does the Markov chain have?
b. Compute the sampling distribution for each variable, conditioned on its Markov blanket. Calculate the **transition matrix Q** containing q(y → y') for all **y, y'**.
Hint: The probability that one of the two sampling variable is chosen is 0.5.
c. What does Q^2, the square of the transition matrix, represent?

a. 2*2=4



P(Rain| Sprinkler=true, WetGrass=true)

Have to sample Cloudy and Rain

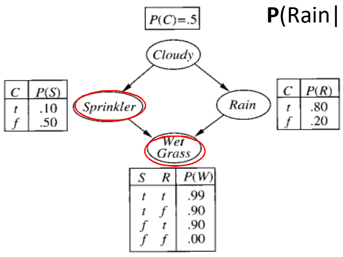
P(C|r,s) = α**P**(C)**P**(s|C)**P**(r|C)= α <0.5,0.5> <0.1,0.5> <0.8,0.2>= α <0.04,0.05>= <4/9,5/9>

P(C|¬r,s) = α**P**(C)**P**(s|C)**P**(¬r|C)=α <0.5,0.5> <0.1,0.5> <0.2,0.8>= α <0.04,0.05>= <1/21,20/21>

P(R|c,s,w) = α**P**(R|c)**P**(w|s,R)=α <0.8,0.2> <0.99,0.9>= α <0.792,0.18>= <22/27,5/27>

P(R|¬c,s,w) = α**P**(R|¬c)**P**(w|s,R)=α <0.2,0.8> <0.99,0.9>= α <0.198,0.72>= <11/51,40/51>

7



P(Rain| Sprinkler=true, WetGrass=true)

Transition matrix?

P(C|r,s) = <4/9,5/9>

P(C|¬r,s) = <1/21,20/21>

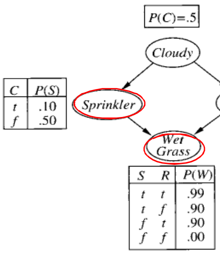
P(R|c,s,w) = <22/27,5/27>

P(R|¬c,s,w) = <11/51,40/51>

Entries where 2 variables are sampled do not exist=0

	(c,r)	$(c,\neg r)$	$(\neg c,r)$	$(\neg c,\neg r)$
(c,r)				0
$(c,\neg r)$			0	
$(\neg c,r)$		0		
$(\neg c,\neg r)$	0			

8



P(Rain| Sprinkler=true, WetGrass=true)

Transition matrix?

P(C|r,s) = <4/9,5/9>

P(C|¬r,s) = <1/21,20/21>

P(R|c,s,w) = <22/27,5/27>

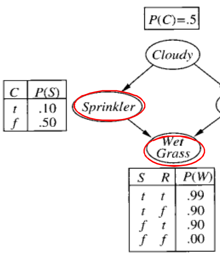
P(R|¬c,s,w) = <11/51,40/51>

Entries on the diagonal are self loops. Transitions can occur by sampling either variable.

Do it for (c,r)→(c,r)= 1/2 P(c|r,s)+1/2 P(r|(c,s,w)= 1/2 (4/9+ 22/27) = 1/2 (12/27+22/27) = 17/27

	(c,r)	$(c,\neg r)$	$(\neg c,r)$	$(\neg c,\neg r)$
(c,r)	17/27			0
$(c,\neg r)$			0	
$(\neg c,r)$		0		
$(\neg c,\neg r)$	0			

9



P(Rain| Sprinkler=true, WetGrass=true)

Transition matrix?

P(C|r,s) = <4/9,5/9>

P(C|¬r,s) = <1/21,20/21>

P(R|c,s,w) = <22/27,5/27>

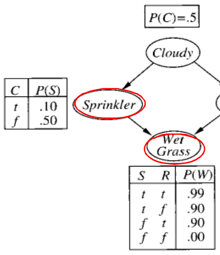
P(R|¬c,s,w) =<11/51,40/51>

Entries where one variable is changed must sample that variable.

Do it for (c,r)→(c,¬r)=1/2 P(¬r|c,s,w)=1/2 * 5/27 = 5/54

	(c,r)	$(c,\neg r)$	$(\neg c,r)$	$(\neg c,\neg r)$
(c,r)	17/27	5/54		0
$(c,\neg r)$			0	
$(\neg c,r)$		0		
$(\neg c,\neg r)$	0			

10



P(Rain | Sprinkler=true, WetGrass=true)

Transition matrix?

P(C | r,s) = <4/9,5/9>

P(C | ¬r,s) = <1/21,20/21>

P(R | c,s,w) = <22/27,5/27>

P(R | ¬c,s,w) =<11/51,40/51>

Entries where one variable is changed must sample that variable.

Do it for (c,r)→(c,¬r)=0.5P(¬r | c,s,w)=1/2 5/27 = 5/54

$$
\begin{array}{ll}
 & \begin{array}{cccc} (c,r) & (c,\neg r) & (\neg c,r) & (\neg c,\neg r) \end{array} \\
\begin{array}{l} (c,r) \\ (c,\neg r) \\ (\neg c,r) \\ (\neg c,\neg r) \end{array} &
\begin{pmatrix}
17/27 & 5/54 & 5/18 & 0 \\
11/27 & 22/189 & 0 & 10/21 \\
2/9 & 0 & 59/153 & 20/51 \\
0 & 1/42 & 11/102 & 310/357
\end{pmatrix}
\end{array}
$$

c. What is Q^2

Q2 represents the probability of going from each state to each state in two steps.

4. Dynamic Bayesian Networks (DBN) can be used to model temporal aspects of the real world. Name and explain assumptions that can be made to reduce the potential complexity of arbitrary DBNs

Stationary Process

The CPTs have the same values in all time slices.

$P(U_t|Parent(U_t))$ is the same for all *t and U*

Markov assumption, Transition Model

The current state depends only on a finite history of previous states. First order:

$$P(X_t|X_{0:t-1}) = P(X_t|X_{t-1})$$

Markov assumption, Sensor Model

The current observation depends only on the current state.

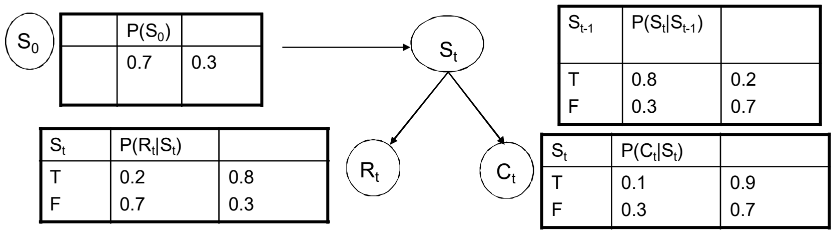
$$P(E_t|X_{0:t}, E_{0:t-1}) = P(E_t|X_t)$$

5. A professor wants to know if students are getting enough sleep. Each day, the professor observes whether the students sleep in class, and whether they have red eyes. The professor has the following domain theory:
   - The prior probability of getting enough sleep, with no observations, is 0.7.
   - The probability of getting enough sleep on night **t** is 0.8 given that the student got enough sleep the previous night, and 0.3 if not.
   - The probability of having red eyes is 0.2 if the student got enough sleep, and 0.7 if not.
   - The probability of sleeping in class is 0.1 if the student got enough sleep, and 0.3 if not.

   Formulate this information as a dynamic Bayesian network that the professor could use to filter or predict from a sequence of observations

   Variables:

   S_0, S_t – enough **s**leep; R_t – **r**ed eyes; C_t – sleep in **c**lass.



	$P(S_0)$	
	0.7	0.3

S_{t-1}	$P(S_t \mid S_{t-1})$	
T	0.8	0.2
F	0.3	0.7

S_t	$P(R_t \mid S_t)$	
T	0.2	0.8
F	0.7	0.3

S_t	$P(C_t \mid S_t)$	
T	0.1	0.9
F	0.3	0.7

6. For the DBN of exercise 5 and for the evidence values

   e_1 = not red eyes, not sleeping in class

   e_2 = red eyes, not sleeping in class

   e_3 = red eyes, sleeping in class

   perform the following computation:

   a. State estimation: Compute P(*EnoughSleep*$_t$ | $e_{1:t}$) for each of t = 1, 2, 3.
   b. Reformulate the DB with only one evidence variable. Give the complete probability tables for the model.

a. State estimation: Compute P(*EnoughSleep*$_t$ | $e_{1:t}$) for t = 1, 2, 3.

$$P(X_{t+1}/e_{1:t+1}) = \alpha P(e_{t+1}/X_{t+1}) \sum_{X_t} \underbrace{P(X_{t+1}/x_t)P(x_t/e_{1:t})}_{P(X_{t+1} \mid e_{1:t})}$$

	$P(S_0)$	
	0.7	0.3

S_{t-1}	$P(S_t \mid S_{t-1})$
T	0.8
F	0.3

S_t	$P(R_t \mid S_t)$
T	0.2
F	0.7

S_t	$P(C_t \mid S_t)$
T	0.1
F	0.3

$P(S_0) =< 0.7, 0.3 >$

$P(S_1|e_1)$

$P(S_1) = \sum_{s_0} P(S_1|s_0)P(s_0) =< 0.8, 0.2 > 0.7 + < 0.3, 0.7 > 0.3 =< 0.65, 0.35 >$

$P(S_1|e_1) = \alpha(\mathrm{P}e_1|\mathrm{S}_1)\mathrm{P}(\mathrm{S}_1) = \alpha < 0.8 \times 0.9, 0.3 \times 0.7 > < 0.65 \times 0.35 >$

$= \alpha < 0.72, 0.21 >< 0.65, 0.3 > =< 0.864, 0.1357>$

e_1 = not red eyes, not sleeping in class

e_2 = red eyes, not sleeping in class

e_3 = red eyes, sleeping in class

a. State estimation: Compute P(*EnoughSleep*$_t$ | $e_{1:t}$) for t = 1, 2, 3.

$$P(X_{t+1}/e_{1:t+1}) = \alpha P(e_{t+1}/X_{t+1}) \sum_{X_t} \underbrace{P(X_{t+1}/x_t)P(x_t/e_{1:t})}_{P(X_{t+1} \mid e_{1:t})}$$

	$P(S_0)$	
	0.7	0.3

S_{t-1}	$P(S_t \mid S_{t-1})$
T	0.8
F	0.3

S_t	$P(R_t \mid S_t)$
T	0.2
F	0.7

S_t	$P(C_t \mid S_t)$
T	0.1
F	0.3

$P(S_1|e_1) =< 0.864, 0.1357>$

$P(\mathrm{S}_2 | \mathrm{e}_{1)} = \sum_{s_1} P(S_2|s_1)P(s_1|e_1) = (< 0.8, 0.2 > \times 0.8643 + < 0.3, 0.7 > \times 0.1357 >)$

$=< 0.7321, 0.2679 >$

$P(S_2|e_{1:2}) = \alpha(\mathrm{P}e_2|\mathrm{S}_2)\mathrm{P}(\mathrm{S}_1|\mathrm{e}_1)$

$= \alpha < 0.2 \times 0.9, 0.7 \times 0.7 >< 0.7321, 0.2679 >$

$= \alpha < 0.18, 0.49 >< 0.7321, 0.2679 >$

$=< 0.510, 0.499 >$

e_1 = not red eyes, not sleeping in class

e_2 = red eyes, not sleeping in class

e_3 = red eyes, sleeping in class

a. State estimation: Compute P(*EnoughSleep*$_t$ | $e_{1:t}$) for t = 1, 2, 3.

$$P(X_{t+1}/e_{1:t+1}) = \alpha P(e_{t+1}/X_{t+1}) \underbrace{\sum_{X_t} P(X_{t+1}/x_t)P(x_t/e_{1:t})}_{P(X_{t+1} \mid e_{1:t})}$$

	$P(S_0)$	
	0.7	0.3

S_{t-1}	$P(S_t \mid S_{t-1})$
T	0.8
F	0.3

S_t	$P(R_t \mid S_t)$
T	0.2
F	0.7

S_t	$P(C_t \mid S_t)$
T	0.1
F	0.3

$P(S_1|e_1) =< 0.864, 0.1357>$

$P(S_2|e_{1:2}) = < 0.510, 0.499 >$

$P(S_3|e_{1:2}) = \sum_{s_2} P(S_3|s_2)P(s_2|e_{1:2}) \quad = (< 0.8, 0.2 > \times 0.5010 + < 0.3, 0.7 > \times 0.4990 >)$

$= < 0.5505, 0.4495 >$

$P(S_3|e_{1:3}) = \alpha P(e_3|S_3)P(S_3|e_{1:2})$

$= \alpha < 0.2 \times 0.1, 0.7 \times 0.3 >< 0.5505, 0.4495 >$

$= \alpha < 0.02, 0.21 >< 0.5505, 0.4495 >$

$= < 0.1045, 0.8955 >$

e_1 = not red eyes, not sleeping in class

e_2 = red eyes, not sleeping in class

e_3 = red eyes, sleeping in class

a. State estimation: Compute P(*EnoughSleep*$_t$ | $e_{1:t}$) for t = 1, 2, 3.

$$P(X_{t+1}/e_{1:t+1}) = \alpha P(e_{t+1}/X_{t+1}) \underbrace{\sum_{X_t} P(X_{t+1}/x_t)P(x_t/e_{1:t})}_{P(X_{t+1} \mid e_{1:t})}$$

	$P(S_0)$	
	0.7	0.3

S_{t-1}	$P(S_t \mid S_{t-1})$
T	0.8
F	0.3

S_t	$P(R_t \mid S_t)$
T	0.2
F	0.7

S_t	$P(C_t \mid S_t)$
T	0.1
F	0.3

$P(S_1|e_1) =< 0.864, 0.1357>$

$P(S_2|e_{1:2}) = < 0.510, 0.499 >$

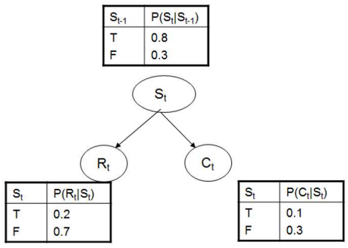
$P(S_3|e_{1:3}) =< 0.1045, 0.8955 >$

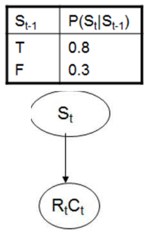
e_1 = not red eyes, not sleeping in class

e_2 = red eyes, not sleeping in class

e_3 = red eyes, sleeping in class

d. Reformulate the DBN as a Hidden Markov Model. Give the complete probability tables for the model.





S_t	P(Rt=true Ct=true|St)	P(Rt=true Ct=false|St)	P(Rt=false Ct=true|St)	P(Rt=false Ct=false|St)
T	0.2*0.1	0.2*0.9	0.8*0.1	0.8*0.9
F	0.7*0.3	0.7*0.7	0.3*0.3	0.3*0.7