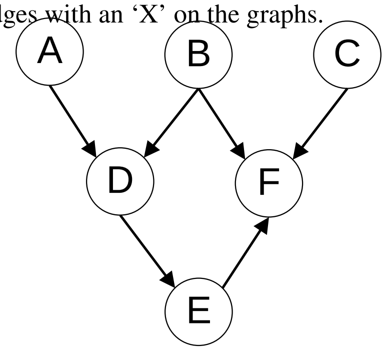
Q1. Bayes Nets

(a) For the following graphs, explicitly state the minimum size set of edges that must be removed such that the corresponding independence relations are guaranteed to be true.

Marked the removed edges with an 'X' on the graphs.

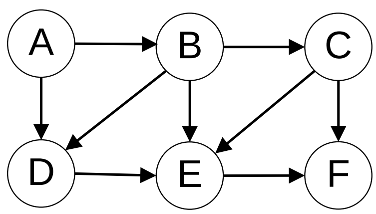


$A \perp\!\!\!\perp B | F$

$A \perp\!\!\!\perp F | D$

$B \perp\!\!\!\perp C$

(i) AD



$A \perp\!\!\!\perp D | B$

$A \perp\!\!\!\perp F | C$

$C \perp\!\!\!\perp D | B$

(ii) AD, (EF OR AB)

(b) You're performing variable elimination over a Bayes Net with variables A, B, C, D, E. So far, you've finished joining over (but not summing out) C, when you realize you've lost the original Bayes Net!

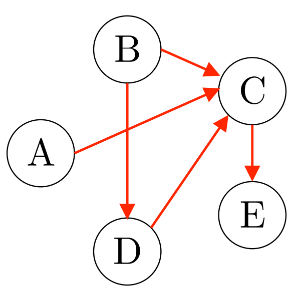
Your current factors are $f(A), f(B), f(B, D), f(A, B, C, D, E)$. Note: these are factors, NOT joint distributions. You don't know which variables are conditioned or unconditioned.

(i) What's the smallest number of edges that could have been in the original Bayes Net? Draw out one such Bayes Net below.

Number of edges = 5

The original Bayes net must have had 5 factors, 1 for each node. f(A) and f(B) must have corresponded to nodes A and B, and indicate that neither A nor B have any parents. f(B, D), then, must correspond to node D, and indicates that D has only B as a parent. Since there is only one factor left, f(A, B, C, D, E), for the nodes C and E, those two nodes must have been joined while you were joining C. This implies two things: 1) E must have had C as a parent, and 2) every other node must have been a parent of either C or E.

The below figure is one possible solution that uses the fewest possible edges to satisfy the above.

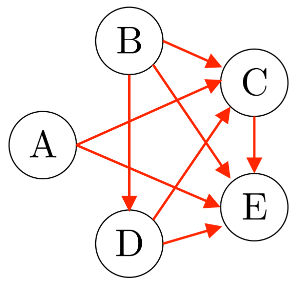


(ii) What's the largest number of edges that could have been in the original Bayes Net? Draw out one such Bayes Net below.

Number of edges = 8

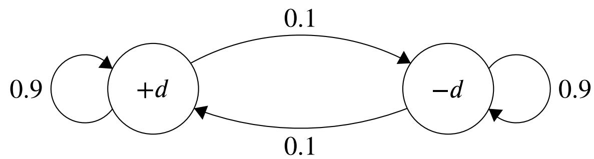
The constraints are the same as outlined in part i). To maximize the number of edges, we make each of A, B, and D a parent of both C and E, as opposed to a parent of one of them.

The below figure is the only possible solution.

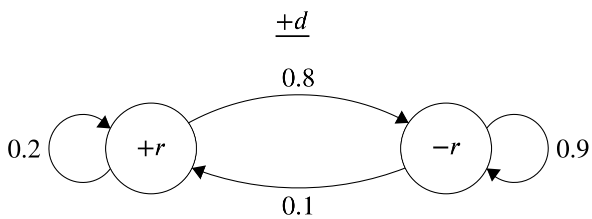


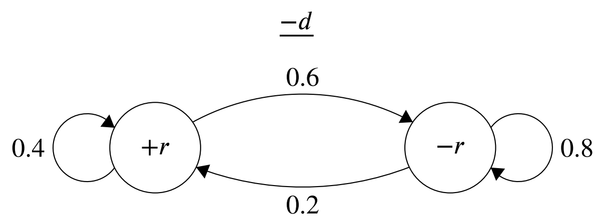
Q2. I Heard You Like Markov Chains

In California, whether it rains or not from each day to the next forms a Markov chain (note: this is a terrible model for real weather). However, sometimes California is in a drought and sometimes it is not. Whether California is in a drought from each day to the next itself forms a Markov chain, and the state of this Markov chain affects the transition probabilities in the rain-or-shine Markov chain. This is the state diagram for droughts:

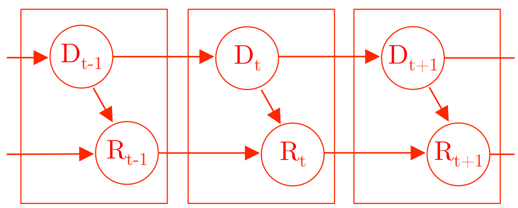


These are the state diagrams for rain given that California is and is not in a drought, respectively:





(a) Draw a dynamic Bayes net which encodes this behavior. Use variables D_{t-1}, D_t, D_{t+1}, R_{t-1}, R_t, and R_{t+1}. Assume that on a given day, it is determined whether or not there is a drought before it is determined whether or not it rains that day.



(b) Draw the CPT for D_t in the above DBN. Fill in the actual numerical probabilities.

$P(D_t \mid D_{t-1})$		
$+d_{t-1}$	$+d_t$	0.9
$+d_{t-1}$	$-d_t$	0.1
$-d_{t-1}$	$+d_t$	0.1
$-d_{t-1}$	$-d_t$	0.9

(c) Draw the CPT for R_t in the above DBN. Fill in the actual numerical probabilities.

$P(R_t \mid R_{t-1}, D_t)$			
$+d_t$	$+r_{t-1}$	$+r_t$	0.2
$+d_t$	$+r_{t-1}$	$-r_t$	0.8
$+d_t$	$-r_{t-1}$	$+r_t$	0.1
$+d_t$	$-r_{t-1}$	$-r_t$	0.9
$-d_t$	$+r_{t-1}$	$+r_t$	0.4
$-d_t$	$+r_{t-1}$	$-r_t$	0.6
$-d_t$	$-r_{t-1}$	$+r_t$	0.2
$-d_t$	$-r_{t-1}$	$-r_t$	0.8

Suppose we are observing the weather on a day-to-day basis, but we cannot directly observe whether California is in a drought or not. We want to predict whether or not it will rain on day $t+1$ given observations of whether or not it rained on days 1 through t.

(d) First, we need to determine whether California will be in a drought on day $t + 1$. Derive a formula for $P(D_{t+1}|r_{1:t})$ in terms of the given probabilities (the transition probabilities on the above state diagrams) and $P(D_t|r_{1:t})$ (that is, you can assume we've already computed the probability there is a drought today given the weather over time).

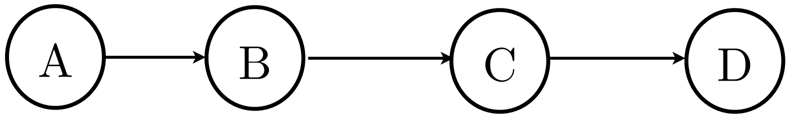
$$P(D_{t+1}|r_{1:t}) = \sum_{d_t} P(D_{t+1}|d_t)P(d_t|r_{1:t})$$

(e) Now derive a formula for $P(R_{t+1}|r_{1:t})$ in terms of $P(D_{t+1}|r_{1:t})$ and the given probabilities.

$$P(R_{t+1}|r_{1:t}) = \sum_{d_{t+1}} P(D_{t+1}|r_{1:t})P(R_{t+1}|r_t, d_{t+1})$$

Q3. Bayes' Nets Sampling

Assume the following Bayes' net, and the corresponding distributions over the variables in the Bayes' net:



$P(A)$	
$-a$	3/4
$+a$	1/4

$P(B\|A)$		
$-a$	$-b$	2/3
$-a$	$+b$	1/3
$+a$	$-b$	4/5
$+a$	$+b$	1/5

$P(C\|B)$		
$-b$	$-c$	1/4
$-b$	$+c$	3/4
$+b$	$-c$	1/2
$+b$	$+c$	1/2

$P(D\|C)$		
$-c$	$-d$	1/8
$-c$	$+d$	7/8
$+c$	$-d$	5/6
$+c$	$+d$	1/6

(a) You are given the following samples:

$+a$	$+b$	$-c$	$-d$
$+a$	$-b$	$+c$	$-d$
$-a$	$+b$	$+c$	$-d$
$-a$	$-b$	$+c$	$-d$

$+a$	$-b$	$-c$	$+d$
$+a$	$+b$	$+c$	$-d$
$-a$	$+b$	$-c$	$+d$
$-a$	$-b$	$+c$	$-d$

(i) Assume that these samples came from performing Prior Sampling, and calculate the sample estimate of $P(+c)$.

5/8

(ii) Now we will estimate $P(+c \mid +a, -d)$. Above, clearly cross out the samples that would **not** be used when doing Rejection Sampling for this task, and write down the sample estimate of $P(+c \mid +a, -d)$ below.

2/3

(b) Using Likelihood Weighting Sampling to estimate $P(-a \mid +b, -d)$, the following samples were obtained. Fill in the weight of each sample in the corresponding row.

Sample				**Weight**
$-a$	$+b$	$+c$	$-d$	$P(+b \mid -a)P(-d \mid +c) = 1/3 * 5/6 = 5/18 = 0.277$
$+a$	$+b$	$+c$	$-d$	$P(+b \mid +a)P(-d \mid +c) = 1/5 * 5/6 = 5/30 = 1/6 = 0.17$
$+a$	$+b$	$-c$	$-d$	$P(+b \mid +a)P(-d \mid -c) = 1/5 * 1/8 = 1/40 = 0.025$
$-a$	$+b$	$-c$	$-d$	$P(+b \mid -a)P(-d \mid -c) = 1/3 * 1/8 = 1/24 = 0.042$

(c) From the weighted samples in the previous question, estimate $P(-a \mid +b, -d)$.

$\frac{5/18+1/24}{5/18+5/30+1/40+1/24} = 0.625$

(d) Which query is better suited for likelihood weighting, $P(D \mid A)$ or $P(A \mid D)$? Justify your answer in one sentence.

$P(D \mid A)$ is better suited for likelihood weighting sampling, because likelihood weighting conditions only on upstream evidence.

(e) Recall that during Gibbs Sampling, samples are generated through an iterative process.

Assume that the only evidence that is available is $A = +a$. Clearly fill in the circle(s) of the sequence(s) below that could have been generated by Gibbs Sampling.

Sequence 1

1 :	$+a$	$-b$	$-c$	$+d$
2 :	$+a$	$-b$	$-c$	$+d$
3 :	$+a$	$-b$	$+c$	$+d$

Sequence 2

1 :	$+a$	$-b$	$-c$	$+d$
2 :	$+a$	$-b$	$-c$	$-d$
3 :	$-a$	$-b$	$-c$	$+d$

Sequence 3

1 :	$+a$	$-b$	$-c$	$+d$
2 :	$+a$	$-b$	$-c$	$-d$
3 :	$+a$	$+b$	$-c$	$-d$

Sequence 4

1 :	$+a$	$-b$	$-c$	$+d$
2 :	$+a$	$-b$	$-c$	$-d$
3 :	$+a$	$+b$	$-c$	$+d$

Gibbs sampling updates one variable at a time and never changes the evidence.

The first and third sequences have at most one variable change per row, and hence could have been generated from Gibbs sampling. In sequence 2, the evidence variable is changed. In sequence 4, the second and third samples have both B and D changing.