CS 188 Summer 2020 | Introduction to Artificial Intelligence | EP 6: BN Solutions

# Q1. Bayes' Nets

| | $P(A)$ |
|---|---|
| $+a$ | 0.25 |
| $-a$ | 0.75 |

| $P(B\|A)$ | $+b$ | $-b$ |
|---|---|---|
| $+a$ | 0.5 | 0.5 |
| $-a$ | 0.25 | 0.75 |

| $P(C\|A)$ | $+c$ | $-c$ |
|---|---|---|
| $+a$ | 0.2 | 0.8 |
| $-a$ | 0.6 | 0.4 |

| $P(D\|B)$ | $+d$ | $-d$ |
|---|---|---|
| $+b$ | 0.6 | 0.4 |
| $-b$ | 0.8 | 0.2 |

| $P(E\|B)$ | $+e$ | $-e$ |
|---|---|---|
| $+b$ | 0.25 | 0.75 |
| $-b$ | 0.1 | 0.9 |

**(a)** Using the Bayes' Net and conditional probability tables above, calculate the following quantities:

**(i)** $P(+a, +b) = 0.25 * 0.5 = 0.125 = \frac{1}{8}$

**(ii)** $P(+a \mid +b) = \frac{0.25*0.5}{0.25*0.5+0.25*0.75} = 0.4 = \frac{2}{5}$

**(iii)** $P(+b \mid +a) = 0.5$

**(b)** Now we are going to consider variable elimination in the Bayes' Net above.

**(i)** Assume we have the evidence $+c$ and wish to calculate $P(E \mid +c)$. What factors do we have initially? $P(A)$, $P(B \mid A)$, $P(+c \mid A)$, $P(D \mid B)$, $P(E \mid B)$

**(ii)** If we eliminate variable B, we create a new factor. What probability does that factor correspond to? $P(D, E \mid A)$

This is the same figure as the previous page, repeated here for your convenience:

| | $P(A)$ |
|---|---|
| $+a$ | 0.25 |
| $-a$ | 0.75 |

| $P(B\|A)$ | $+b$ | $-b$ |
|---|---|---|
| $+a$ | 0.5 | 0.5 |
| $-a$ | 0.25 | 0.75 |

| $P(C\|A)$ | $+c$ | $-c$ |
|---|---|---|
| $+a$ | 0.2 | 0.8 |
| $-a$ | 0.6 | 0.4 |

| $P(D\|B)$ | $+d$ | $-d$ |
|---|---|---|
| $+b$ | 0.6 | 0.4 |
| $-b$ | 0.8 | 0.2 |

| $P(E\|B)$ | $+e$ | $-e$ |
|---|---|---|
| $+b$ | 0.25 | 0.75 |
| $-b$ | 0.1 | 0.9 |

**(iii)** What is the equation to calculate the factor we create when eliminating variable B? $f(A, D, E) = \sum_B P(B \mid A) * P(D \mid B) * P(E \mid B)$

**(iv)** After eliminating variable B, what are the new set of factors? As in (ii), write the probabilities that the factors represent. For each factor, also provide its size. Use only as many rows as you need to.

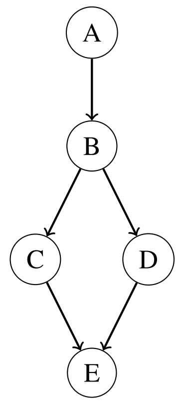
| Factor | Size after elimination |
|---|---|
| $P(A)$ | 2 |
| $P(+\text{c} \mid A)$ | 2 |
| $P(D, E \mid A)$ | $2^3$ |
| | |

**(v)** Now assume we have the evidence $-c$ and are trying to calculate $P(A|-c)$. What is the most efficient elimination ordering? If more than one ordering is most efficient, provide any one of them. E, D, B or D, E, B

**(vi)** Once we have run variable elimination and have $f(A, -c)$ how do we calculate $P(+a \mid -c)$? (give an equation) $\frac{f(+a,-c)}{f(+a,-c)+f(-a,-c)}$ or note that elimination is unnecessary - just use Bayes' rule

# Q2. Bayes' Nets: Sampling

Assume we are given the following Bayes' net, with the associated conditional probability tables (CPTs).




| A | P(A) |
|---|---|
| $+a$ | 0.5 |
| $-a$ | 0.5 |

| A | B | P(B \| A) |
|---|---|---|
| $+a$ | $+b$ | 0.2 |
| $+a$ | $-b$ | 0.8 |
| $-a$ | $+b$ | 0.5 |
| $-a$ | $-b$ | 0.5 |

| B | C | P(C \| B) |
|---|---|---|
| $+b$ | $+c$ | 0.4 |
| $+b$ | $-c$ | 0.6 |
| $-b$ | $+c$ | 0.8 |
| $-b$ | $-c$ | 0.2 |

| B | D | P(D \| B) |
|---|---|---|
| $+b$ | $+d$ | 0.2 |
| $+b$ | $-d$ | 0.8 |
| $-b$ | $+d$ | 0.2 |
| $-b$ | $-d$ | 0.8 |

| C | D | E | P(E \| C, D) |
|---|---|---|---|
| $+c$ | $+d$ | $+e$ | 0.6 |
| $+c$ | $+d$ | $-e$ | 0.4 |
| $+c$ | $-d$ | $+e$ | 0.2 |
| $+c$ | $-d$ | $-e$ | 0.8 |
| $-c$ | $+d$ | $+e$ | 0.4 |
| $-c$ | $+d$ | $-e$ | 0.6 |
| $-c$ | $-d$ | $+e$ | 0.8 |
| $-c$ | $-d$ | $-e$ | 0.2 |

You are given a set of the following samples, but are not told whether they were collected with rejection sampling or likelihood weighting.

| | | | | |
|---|---|---|---|---|
| $-a$ | $-b$ | $+c$ | $+d$ | $+e$ |
| $-a$ | $+b$ | $+c$ | $-d$ | $+e$ |
| $-a$ | $-b$ | $-c$ | $-d$ | $+e$ |
| $-a$ | $-b$ | $+c$ | $-d$ | $+e$ |
| $-a$ | $+b$ | $+c$ | $+d$ | $+e$ |

Throughout this problem, you may answer as either numeric expressions (e.g. $0.1*0.5$) or numeric values (e.g. 0.05).

**(a)** Assuming these samples were generated from *rejection sampling*, what is the sample based estimate of $P(+b \mid -a, +e)$?

Answer: 0.4

The answer is the number of samples satisfying the query variable's assignment (in this case, $B = +b$ divided by the total number of samples, so the answer is 2 / 5 = 0.4.

**(b)** Assuming these samples were generated from *likelihood weighting*, what is the sample-based estimate of $P(+b \mid -a, +e)$?

Answer: $\frac{1}{3}$

Based on likelihood weighting, we know the weight of each sample is $P(A = a) * P(E = e \mid C = c, D = d)$. The weights are: 0.3 (= 0.5 * 0.6), 0.1 (= 0.5 * 0.2), 0.4 (= 0.5 * 0.8), 0.1 (same assignments to $C$ and $D$ as second sample), 0.3 (same assignments to $C$ and $D$ as first sample). The estimate is then (0.1 + 0.3) / (0.3 + 0.1 + 0.4 + 0.1 + 0.3) = 0.4 / 1.20 = 1/3 = 0.333.

**(c)** Again, assume these samples were generated from *likelihood weighting*. However, you are not sure about the original CPT for $P(E \mid C, D)$ given above being the CPT associated with the Bayes' Net: With 50% chance, the CPT associated with the Bayes' Net is the original one. With the other 50% chance, the CPT is actually the CPT below.

| C | D | E | P(E \| C, D) |
|---|---|---|---|
| $+c$ | $+d$ | $+e$ | 0.8 |
| $+c$ | $+d$ | $-e$ | 0.2 |
| $+c$ | $-d$ | $+e$ | 0.4 |
| $+c$ | $-d$ | $-e$ | 0.6 |
| $-c$ | $+d$ | $+e$ | 0.2 |
| $-c$ | $+d$ | $-e$ | 0.8 |
| $-c$ | $-d$ | $+e$ | 0.6 |
| $-c$ | $-d$ | $-e$ | 0.4 |

Samples from previous page copied below for convenience:

| | | | | |
|---|---|---|---|---|
| $-a$ | $-b$ | $+c$ | $+d$ | $+e$ |
| $-a$ | $+b$ | $+c$ | $-d$ | $+e$ |
| $-a$ | $-b$ | $-c$ | $-d$ | $+e$ |
| $-a$ | $-b$ | $+c$ | $-d$ | $+e$ |
| $-a$ | $+b$ | $+c$ | $+d$ | $+e$ |

Given this uncertainty, what is the sample-based estimate of $P(+b \mid -a, +e)$?

Answer: $\frac{10}{27}$

The weight of each sample is $P(A = a) * (0.5 * P_1(E = e \mid C = c, D = d) + 0.5 * P_2(E = e \mid C = c, D = d))$. The new weights are 0.35 (= 0.5 * (0.5 * 0.6 + 0.5 * 0.8)), 0.15 (= 0.5 * (0.5 * 0.2 + 0.5 * 0.4)), 0.35 (= 0.5 * (0.5 * 0.8 + 0.5 * 0.6)), 0.15, and 0.35. The estimate is then (0.15 + 0.35) / (0.35 * 3 + 0.15 * 2) = 0.5 / 1.35 = 10 / 27

**(d)** Now assume you can only sample a *small, limited number of samples*, and you want to estimate $P(+b, +d \mid -a)$ and $P(+b, +d \mid +e)$. You are allowed to estimate the answer to one query with likelihood weighting, and the other answer with rejection sampling. In order to obtain the best estimates for both queries, *which query should you estimate with likelihood weighting*? (The other query will have to be estimated with rejection sampling.)

- ○ $P(+b, +d \mid -a)$
- ● $P(+b, +d \mid +e)$
- ○ Either – both choices allow you to obtain the best estimates for both queries.

The evidence +e is at the leaves of the Bayes' net, which means it's possible to sample all the other variables, but have to reject the last node E. We can avoid this problem by using likelihood weighting for sampling, since it fixes the values of observed random variables to that of the fixed evidence.

**(e)** Suppose you choose to use Gibbs sampling to estimate $P(B, E \mid +c, -d)$. Assume the CPTs are the same as the ones for parts (a) and (b). Currently your assignments are the following:

$-a \quad -b \quad +c \quad -d \quad +e$

**(i)** Suppose the next step is to resample E.
What is the probability that the new assignment to E will be +e?

Answer: 0.2 In order to sample $E$, we need to calculate $P(E \mid -a, -b, +c, -d)$, which is equal to $P(E \mid +c, -d)$ since E is conditionally independent of A and B, given C and D. The value for $P(+e \mid +c, -d)$ is given directly in the CPT for $P(E|C, D)$, and it is 0.2.

**(ii)** Instead, suppose the next step is to resample $A$.
What is the probability that the new assignment to $A$ will be $+a$?

Answer: $\frac{8}{13}$ In order to sample $A$, we need to calculate $P(A \mid -b, +c, -d, +e)$, which is equal to $P(A \mid B)$ since A is conditionally independent of C, D, and E, given B. We can calculate this using Bayes' rule: $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$. Thus, $P(+a \mid -b) = \frac{P(-b|+a)P(+a)}{P(-b)} = \frac{P(-b|+a)P(+a)}{\sum_{a\in\{+a,-a\}} P(-b|a)}P(a) = \frac{0.8*0.5}{0.8*0.5+0.5*0.5)} = \frac{0.4}{0.65} = \frac{8}{13}$

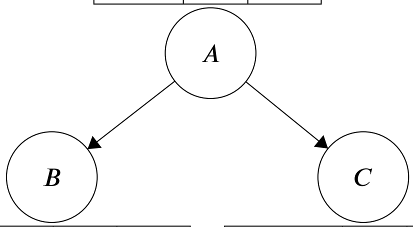
**(iii)** Instead, suppose the next step is to resample B.
What is the probability that the new assignment to B will be +b?

Answer: $\frac{1}{3}$ In order to sample $B$, we need to calculate $P(B \mid -a, +c, -d, +e)$, which is equal to $P(B \mid -a, +c, -d)$ since B is conditionally independent of E, given C and D. The CPT tables that are involved in calculating $P(B \mid -a + c, -d, +e)$ are $P(A)$, $P(B|A)$, $P(C|B)$, $P(D|B)$. First, we remove rows of the CPTs that do not agree with the evidence -a, +c, -d, +e. We then join the resulting CPTs to obtain $P(-a, B, +c, -d)$. We select $P(-a, +b, +c, -d)$ and $P(-a, -b, +c, -d)$ from this table, and normalize so that the two probabilities sum to one (i.e., to transform them to $P(+b| - a, +c, -d)$ and $P(-b| - a, +c, -d)$.

# Q3. Sampling

Consider the following Bayes net. The joint distribution is not given, but it may be helpful to fill in the table before answering the following questions.

| | | |
|---|---|---|
| $P(A)$ | $+a$ | $2/3$ |
| | $-a$ | $1/3$ |




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

| | | | |
|---|---|---|---|
| $P(C\|A)$ | $+a$ | $+c$ | $1/2$ |
| | $+a$ | $-c$ | $1/2$ |
| | $-a$ | $+c$ | $2/3$ |
| | $-a$ | $-c$ | $1/3$ |

| $P(A,B,C)$ | | | |
|---|---|---|---|
| $+a$ | $+b$ | $+c$ | 1/6 |
| $+a$ | $+b$ | $-c$ | 1/6 |
| $+a$ | $-b$ | $+c$ | 1/6 |
| $+a$ | $-b$ | $-c$ | 1/6 |
| $-a$ | $+b$ | $+c$ | 1/18 |
| $-a$ | $+b$ | $-c$ | 1/36 |
| $-a$ | $-b$ | $+c$ | 1/6 |
| $-a$ | $-b$ | $-c$ | 1/12 |

We are going to use sampling to approximate the query $P(C|+b)$. Consider the following samples:

| Sample 1 | Sample 2 | Sample 3 |
|---|---|---|
| $(+a,+b,+c)$ | $(+a,-b,-c)$ | $(-a,+b,+c)$ |

**(a)** Fill in the following table with the probabilities of drawing each respective sample given that we are using each of the following sampling techniques.

$P(+b) = \frac{2}{6} + \frac{1}{12} = \frac{5}{12}$

| $P$(sample \| method) | Sample 1 | Sample 2 |
|---|---|---|
| Prior Sampling | $1/6$ | $1/6$ |
| Rejection Sampling | $\frac{1/6}{5/12} = 2/5$ | $0$ |
| Likelihood Weighting | $2/3 \cdot 1/2 = 1/3$ | $0$ |

Lastly, we want to figure out the probability of getting Sample 3 by Gibbs sampling. We'll initialize the sample to $(+a,+b,+c)$, and resample $A$ then $C$.

**(b)** What is the probability the sample equals $(-a,+b,+c)$ after resampling $A$?

$P(-a|+b,+c) = \frac{P(-a,+b,+c)}{P(-a,+b,+c)+P(+a,+b,+c)} = \frac{1/18}{1/18+1/6} = \frac{1/18}{4/18} = \frac{1}{4}$

**(c)** What is the probability the sample equals $(-a,+b,+c)$ after resampling $C$, given that the sample equals $(-a,+b,+c)$ after resampling $A$?

$P(+c|-a,+b) = P(+c|-a) = \frac{2}{3}$

**(d)** What is the probability of drawing Sample 3, $(-a,+b,+c)$, using Gibbs sampling in this way?

$P(-a|+b,+c) \cdot P(+c|-a,+b) = \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6}$

**(e)** Suppose that through some sort of accident, we lost the probability tables associated with this Bayes net. We recognize that the Bayes net has the same form as a naïve Bayes problem. Given our three samples:

$$(+a, +b, +c), (+a, -b, -c), (-a, +b, +c)$$

Use naïve Bayes maximum likelihood estimation to approximate the parameters in all three probability tables.

| | | |
|---|---|---|
| $P(A)$ | $+a$ | $2/3$ |
| | $-a$ | $1/3$ |

| | | | |
|---|---|---|---|
| $P(B\|A)$ | $+a$ | $+b$ | $1/2$ |
| | $+a$ | $-b$ | $1/2$ |
| | $-a$ | $+b$ | $1$ |
| | $-a$ | $-b$ | $0$ |

| | | | |
|---|---|---|---|
| $P(C\|A)$ | $+a$ | $+c$ | $1/2$ |
| | $+a$ | $-c$ | $1/2$ |
| | $-a$ | $+c$ | $1$ |
| | $-a$ | $-c$ | $0$ |

**(f)** What problem would Laplace smoothing fix with the maximum likelihood estimation parameters above?

Laplace smoothing would help prevent overfitting to our very few number of samples. It would avoid the zero probabilities found in the parameters above. It would bring the estimated parameters closer to uniform, which in this case is closer to the original parameters than the maximum likelihood estimated parameters.