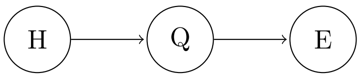
# 1 Conceptual Review

1. When would we want to use inference?
   Inference is used to calculate some useful quantity from a joint probability distribution. For example, we can use it in speech recognition (the example given in class). Here, we could answer a question such as what is the most probable next word given the audio for the next word and the fact that the first word is "artificial". In this case, inference tells us which words were most likely to be said in the audio clip.

2. Suppose we are given binary random variables Q, H, E (query, hidden, evidence).
   We want to query $P(q \mid e)$.




   (a) **Enumeration**
   Perform inference on a joint distribution. Use the Bayes net above to break down joint into CPT factors.

   *Note:* You may use a proportionality constant $\alpha$ in your answer.

   Inference on a joint distribution:
   $P(q \mid e) = \alpha P(q, e)$

   $= \alpha \sum_{h \in (h_1, h_2)} P(q, h, e)$

   Using Bayes net to break down joint in to CPT factors:
   $P(q \mid e) = \alpha \sum_{h \in (h_1, h_2)} P(h)P(q \mid h)P(e \mid q)$

   $= \alpha[P(h_1)P(q \mid h_1)P(e \mid q) + P(h_2)P(q \mid h_2)P(e \mid q)]$

   (b) **Variable Elimination**
   Rewrite your answer to enumeration by moving summations inwards as far as possible.

   *Note:* You may use a proportionality constant $\alpha$ in your answer.

   $P(q \mid e) = \alpha P(e \mid q) \sum_{h \in (h_1, h_2)} P(h)P(q \mid h)$

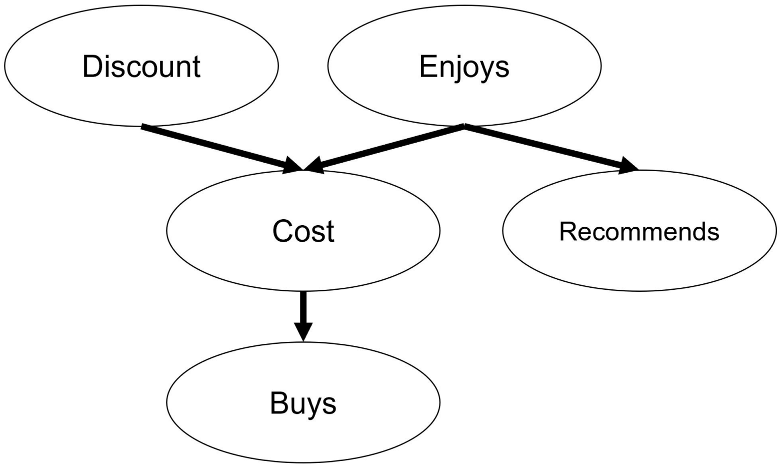
   $= \alpha P(e \mid q)[P(h_1)P(q \mid h_1) + P(h_2)P(q \mid h_2)]$

   (c) Based on 2a and 2b, why is variable elimination more efficient than enumeration?
   We sum entries from the joint distribution (these entries are obtained from a Bayes Net by multiplying conditional probabilities) in order to do inference by enumeration. This involves multiple repeated sub-expressions. We can move the summations inwards as far as possible in order to eliminate repeated computation. This is called variable elimination and is thus more efficient.

## 2 Inference

Realizing that students aren't particularly fond of reading the textbook, the 281 course staff have developed a software that automatically scans the textbook and outputs key points for each individual chapter. However, since the development of the software requires time and computational resources, the 281 staff decides to offer a free one month trial to students, after which a paid subscription is necessary to keep using the software. The following network and variables are used to represent the problem:

- $Discount(D)$: $+d$ if a discount is offered, $-d$ otherwise
- $Enjoys(E)$: $+e$ if a student enjoys the software, $-e$ otherwise
- $Cost(C)$: $+c$ if the software cost is $< 20$, $-c$ otherwise
- $Recommends(R)$: $+s$ if the student recommends the software to a friend, $-s$ otherwise
- $Buys(B)$: $+b$ if the student buys a software subscription, $-b$ otherwise




1. How can we represent the probability that a student buys and recommends the software using the conditional probabilities at each node?
   $P(+b,+r) = \sum_{c,d,e} P(+b|c)P(c|d,e)P(d)P(e)P(+r|e)$

   This sum is equivalent to summing out the hidden variables in the joint distribution: $\sum_{c,d,e} P(d,e,c,+r,+b)$.
2. The staff has surveyed students and collected data on whether the students enjoyed the software or not. With this information, we want to perform inference on a joint distribution where the query variable is $Buys$ $(B)$.
   (a) How can we represent the probability expression in terms of conditional probabilities from the network?
   $P(B|E) = \alpha P(B,E) = \alpha \sum_{d,c,r} P(B,E,d,c,r) = P(E)\alpha \sum_{d,c,r} P(d)P(c|d,E)P(B|c)P(r|E)$

   Note: Equation 13.9 on page 493 of the TB goes into detail about why we use $\alpha$. In short, when we are calculating conditional probabilities, $\alpha$ acts as a normalization constant. However, we can proceed with calculating the conditional probabilities even without knowing the value of $\alpha$ because relative proportions remain the same without normalization (e.g. relative proportions of $P(+b|E)$ and $P(-b|E)$ remain the same without knowing the exact value of $\alpha = 1/P(E)$).
   (b) What are the hidden and evidence variable(s)?
   The hidden variables are $D, C, R$, and the evidence variable is $E$.
3. Using the probability expression from the previous part, we want to compute the query $B$ given evidence that the student enjoys the software. Assume the variable ordering is in alphabetical order.
   (a) How many factors are there, and what are the dimensions of each factor?
   Our expression is: $P(B|+e) = \alpha P(+e) \sum_{r,d,c} P(d)P(c|d,+e)P(B|c)P(r|+e)$

   $= \alpha P(+e) \sum_r P(r|+e) \sum_d P(d) \sum_c P(c|d,+e)(B|c)$

   Each conditional probability corresponds to an individual factor, so there are 5 factors total.

The factor for $P(d)$ and $P(r|+e)$ each have dimension $2 \times 1$, the factors for $P(c|d,+e)$ and $P(B|c)$ each have dimension $2 \times 2$, and the factor for $P(+e)$ is a one-element vector.

(b) Run the variable elimination algorithm to eliminate repeated computations for the expression $P(B|+e)$.

All factors: $P(D), P(+e), P(C|D,+e), P(B|C), P(R|+e)$

- Choose C: The relevant factors are $P(C|D,+e), P(B|C)$. We sum out $C$ to get $f_1(D,B) = \sum_c P(C=c|D,+e)P(B|C=c)$.

  Expression: $P(B|+e) = \alpha P(+e) \sum_{d,r} P(D=d)P(R=r|+e) \times f_1(D,B)$
- Choose D: We sum out the relevant factors $P(D)$ and $f_1(D,B)$ to get $f_2(B) = \sum_d f_1(D=d,B)P(d)$.

  Expression: $P(B|+e) = \alpha P(+e) \sum_r P(R=r|+e) \times f_2(B)$
- Choose R: We sum out the relevant factor $P(R|+e)$ to get $\sum_r P(R=r|+e) = 1$. We discard this variable since it is irrelevant and no other factors depend on it.

  Expression: $P(B|+e) = \alpha P(+e) \times f_2(B)$

(c) How does the resulting expression change if the variable ordering is instead in reverse alphabetical order?

- Choose R: We sum out relevant factor $P(R|+e)$ to get $\sum_r P(R=r|+e) = 1$. We can discard this variable since it is irrelevant.
- Choose D: We sum out relevant factors $P(D)$, $P(C|D,+e)$ to get $f_2(C) = \sum_d P(C|D=d,+e) \times P(D=d)$.

  Expression: $P(B|+e) = \alpha P(+e) \sum_c P(B|c) \times f_2(C=c)$
- Choose C: We sum out relevant factors $P(B|c)$ and $f_2(C)$ to get $f_3(B) = \sum_c P(B|C=c) \times f_2(C=c)$.

  $P(B|+e) = \alpha P(+e) f_3(B)$

(d) How do the two orderings compare with respect to time and space complexity?

When the terms were ordered in alphabetical order, the largest factor had 2 variables. When the terms were ordered in reverse alphabetical order, the largest factor had 1 variable. Since the size of the largest factor determines the space/time complexity, the second ordering performs better.

(e) Describe a heuristic that could be useful in determining a variable ordering to minimize the size of the largest factor.

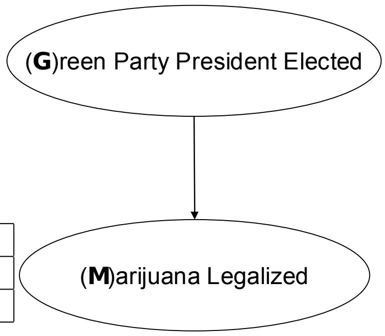
Potential ideas:

- Eliminate whichever variable minimizes the size of the next factor to be constructed.
- Eliminate the variable with the fewest dependent variables

# 3 Bayes' Nets: Green Party President

It's election year again! In a parallel universe the Green Party is running for presidency. Pundits believe that Green Party presidents are more likely to legalize marijuana than candidates from other parties, but legalization could occur under any administration. Armed with the power of probability, the analysts model the situation with the Bayes Net below.

| | +g | -g |
|---|---|---|
| P(G) | 0.1 | 0.9 |




| | P(+m \| G) | P(-m \| G) |
|---|---|---|
| +g | 0.667 | 0.333 |
| -g | 0.25 | 0.75 |

1. Fill in the joint probability table over G and M.

| G | M | P(G, M) |
|---|---|---|
| +g | +m | |
| +g | -m | |
| -g | +m | |
| -g | -m | |

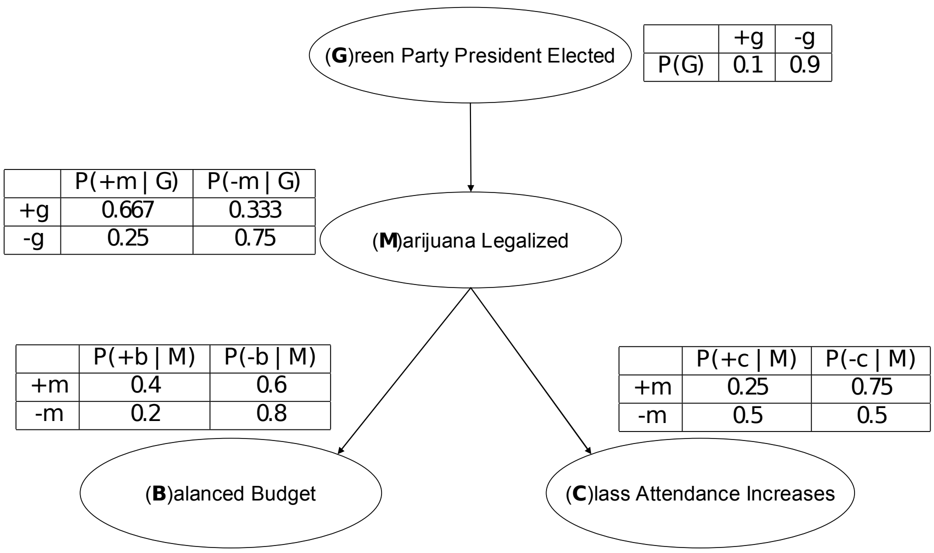
| G | M | P(G, M) |
|---|---|---|
| +g | +m | $1/15$ |
| +g | -m | $1/30$ |
| -g | +m | $9/40$ |
| -g | -m | $27/40$ |

2. What is P(+m), the marginal probability that marijuana is legalized?
$P(+m) = P(+m, +g) + P(+m, -g) = P(+m \mid +g)P(+g) + P(+m \mid -g)P(-g) = \frac{2}{3} \cdot \frac{1}{10} + \frac{1}{4} \cdot \frac{9}{10} = \frac{7}{24}$

3. News agencies air 24/7 coverage of the recent legalization of marijuana (+m), but you can't seem to find out who won the election. What is the conditional probability P(+g | +m) that a Green Party president was elected?

$$P(+g \mid +m) = \frac{P(+g, +m)}{P(+m)} = \frac{P(+m \mid +g)P(+g)}{P(+m)} = \frac{\frac{2}{3} \cdot \frac{1}{10}}{\frac{7}{24}} = \frac{8}{35}$$

We can make better inferences if we observe more evidence. On the next page, we will expand on the model (Bayes net) by introducing two new random variables: whether the budget is balanced (B), and whether class attendance increases (C). The expanded Bayes net and conditional distributions are shown below.




| | +g | -g |
|---|---|---|
| P(G) | 0.1 | 0.9 |

| | P(+m \| G) | P(-m \| G) |
|---|---|---|
| +g | 0.667 | 0.333 |
| -g | 0.25 | 0.75 |

| | P(+b \| M) | P(-b \| M) |
|---|---|---|
| +m | 0.4 | 0.6 |
| -m | 0.2 | 0.8 |

| | P(+c \| M) | P(-c \| M) |
|---|---|---|
| +m | 0.25 | 0.75 |
| -m | 0.5 | 0.5 |

4. The full joint distribution is given below. Fill in the missing values.

| $G$ | $M$ | $B$ | $C$ | $P(G,M,B,C)$ | $G$ | $M$ | $B$ | $C$ | $P(G,M,B,C)$ |
|---|---|---|---|---|---|---|---|---|---|
| + | + | + | + | $1/150$ | - | + | + | + | $9/400$ |
| + | + | + | - | | - | + | + | - | $27/400$ |
| + | + | - | + | $1/100$ | - | + | - | + | $27/800$ |
| + | + | - | - | $3/100$ | - | + | - | - | $81/800$ |
| + | - | + | + | $1/300$ | - | - | + | + | $27/400$ |
| + | - | + | - | $1/300$ | - | - | + | - | $27/400$ |
| + | - | - | + | | - | - | - | + | |
| + | - | - | - | $1/75$ | - | - | - | - | $27/100$ |

| $G$ | $M$ | $B$ | $C$ | $P(G,M,B,C)$ | $G$ | $M$ | $B$ | $C$ | $P(G,M,B,C)$ |
|---|---|---|---|---|---|---|---|---|---|
| + | + | + | + | $1/150$ | - | + | + | + | $9/400$ |
| + | + | + | - | $1/50$ | - | + | + | - | $27/400$ |
| + | + | - | + | $1/100$ | - | + | - | + | $27/800$ |
| + | + | - | - | $3/100$ | - | + | - | - | $81/800$ |
| + | - | + | + | $1/300$ | - | - | + | + | $27/400$ |
| + | - | + | - | $1/300$ | - | - | + | - | $27/400$ |
| + | - | - | + | $1/75$ | - | - | - | + | $27/100$ |
| + | - | - | - | $1/75$ | - | - | - | - | $27/100$ |

5. Compute the following quantities. You may use either the full joint distribution or the conditional tables, whichever is more convenient.

(a) P(+b | +m) =

$\frac{4}{10}$ (directly from conditional)

(b) P(+b | +m, +g) =

$\frac{4}{10}$ (also directly from conditional, since $B \perp\!\!\!\perp G|M$)

(c) P(+b) =

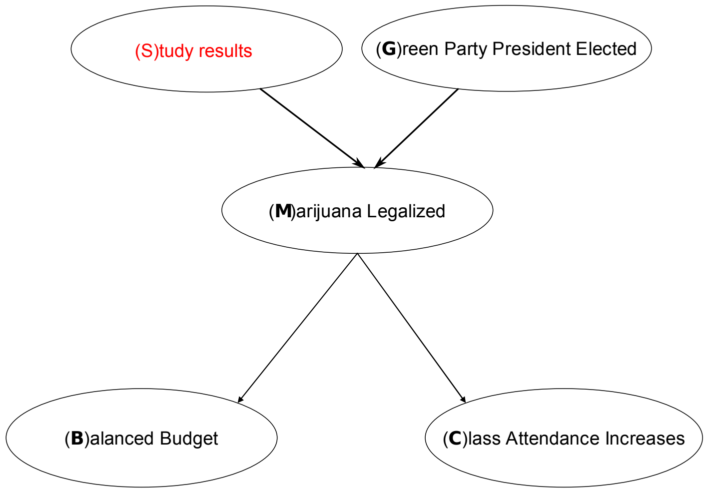
$\sum_{g,m,c} \text{P(g,m,+b,c)} = \frac{8}{1200} + \frac{24}{1200} + \frac{4}{1200} + \frac{4}{1200} + \frac{27}{1200} + \frac{81}{1200} + \frac{81}{1200} + \frac{81}{1200} = \frac{31}{120}$

(summed from joint)

(d) P(+c | +b) =

$\frac{P(+b,\ +c)}{P(+b)} = \frac{\sum_{g,m} P(g,m,+b,+c)}{31/120} = \left(\frac{8}{1200} + \frac{4}{1200} + \frac{27}{1200} + \frac{81}{1200}\right) \cdot \frac{120}{31} = \frac{12}{31}$
(summed from joint)

6. Now, add a node $S$ to the Bayes net above that reflects the possibility that a new scientific study could influence the probability that marijuana is legalized. Assume that the study does not directly influence B or C. Which CPT(s) need to be modified?




$P(M|G)$ will become $P(M|G,S)$, and will contain 8 entries instead of 4.

7. Consider your augmented model. Just based on the structure, which of the following are guaranteed to be true?

(a) $B \perp\!\!\!\perp G$
(b) $C \perp\!\!\!\perp G \mid M$
(c) $G \perp\!\!\!\perp S$
(d) $G \perp\!\!\!\perp S \mid M$
(e) $G \perp\!\!\!\perp S \mid B$
(f) $B \perp\!\!\!\perp C$
(g) $B \perp\!\!\!\perp C \mid G$

(b), (c)