Massachusetts Institute of Technology

6.034 Artificial Intelligence

Solutions # $\boxed{8}$ 6034 Item # $\boxed{}$

Problem 1 Warmup

If we let L denote "left-handed" and T "test is positive", then "10% of the people are left-handed" means $P(L) = 0.10$ and $P(\bar{L}) = 0.90$, and "the test correctly determines your handedness 85% of the time" means $P(T \mid L) = 0.85$ and $P(\bar{T} \mid \bar{L}) = 0.85$.

To compute $P(L \mid T)$, we use Bayes' theorem:

$$\frac{P(L \mid T) = P(L)P(T \mid L)}{P(L)P(T \mid L) + P(\bar{L})P(T \mid \bar{L}),}$$

so:

$$P(L \mid T) = \frac{(0.10)(0.85)}{(0.10)(0.85) + (1 - 0.10)(1 - 0.85)} = 0.386.$$

Problem 2 Conditional Probability

You were asked to compute $P(E = \texttt{\#f} \mid A = \texttt{\#t})$ by hand and to verify that it yields the same answer as

```
(conditional-probability mcbn1 '((E #f)) '((A #t)))
```

The cookbook approach for computing this value would be to write the required conditional probability in terms of a ratio of joint probabilities:

$$P(E = \texttt{\#f} \mid A = \texttt{\#t}) = \frac{P(E = \texttt{\#f}, A = \texttt{\#t})}{P(A = \texttt{\#t})}$$

and then to compute both numerator and denominator. This is, indeed, how our program does it:

```
(define (conditional-probability bnet set on)
  (/ (infer bnet (append set on))
     (infer bnet on)))
```

To compute the numerator by hand, for example, we need to sum the probability of all those circumstances (different sets of values for the five variables) in which $E = \texttt{\#f}$ and $A = \texttt{\#t}$. These will correspond to the eight possible sets of values of the "don't care" variables, B, C, and D. For people, however, this is tedious!

Instead, we may realize that the only "don't care" variable we care about is C, and we can claim that

$$\begin{aligned} P(E = \texttt{\#f} \mid A = \texttt{\#t}) &= P(E = \texttt{\#f} \mid C = \texttt{\#t}) P(C = \texttt{\#t} \mid A = \texttt{\#t}) \\ &\quad + P(E = \texttt{\#f} \mid C = \texttt{\#f}) P(C = \texttt{\#f} \mid A = \texttt{\#t}) \end{aligned}$$

by reading the relationship directly off the graph. This gives us

$$P(E = \texttt{\#f} \mid A = \texttt{\#t}) = (0.9)(0.4) + (0.6)(0.6) = 0.72$$

Breaking down the conditional probability of some state of E given some state of A relies on three observations:

- Conditional probabilities chain when there is conditional independence. Thus, for example,
$$P(\bar{e} \mid a, c) = P(\bar{e} \mid c) P(c \mid a)$$
where, as explained in the problem set, we use a as an abbreviation for $A = \texttt{\#t}$ and $\bar{a}$ for $A = \texttt{\#f}$.
- If E is in fact to be absent in the presence of A, we have to consider each possible value of the intermediate variable C, and sum over the two cases.
- By the conditional independence implied by the `MCBN1` diagram between B and C given A and between D and E given C, the values of the variables B and D cannot affect the relationship we are interested in.

With the program to compute conditional probabilities given above, we get:

```
(conditional-probability mcbn1 '((E #f)) '((A #t)))
=> .72

(conditional-probability mcbn1 '((E #f)) '((B #t)))
=> .7527272727272727

(conditional-probability mcbn1 '((D #t) (E #t)) '((A #t) (B #f)))
=> .076

(conditional-probability mcbn1 '((A #t)) '((D #t) (E #t)))
=> .4517167381974248
```

Problem 3 Independence

Part A

Our solution here uses the `c-independent?` procedure defined later in the problem set:

```
(define (independent? bnet var1 var2)
  (c-independent? bnet var1 var2 '()))

(independent? mcbn1 'A 'B)
=> ()

(independent? mcbn1 'D 'E)
=> ()

(independent? mcbn1 'A 'D)
=> ()
```

In the MCBN1 network, it turns out that no two variables are (unconditionally) independent.

Part B

In the special case that none of the nodes in a network have been observed to be present or absent, then two nodes are independent if they have no common ancestors in the network. A node is, for this, considered an ancestor of itself, so in a network where `X ---> Y`, X is the common ancestor.

Problem 4 Conditional Independence

Part A

The missing information needed to define the problem is $P(\text{raining})$. If we say that $P(\text{raining}) = 0.01$, then we can fill in the rest of the conditional probabilities needed to define the Bayes net.

We know that $P(p) = P(p \mid r)P(r) + P(p \mid \bar{r})P(\bar{r})$. Thus, we can solve

$$0.1 = (0.5)(0.01) + P(p \mid \bar{r})(0.99)$$

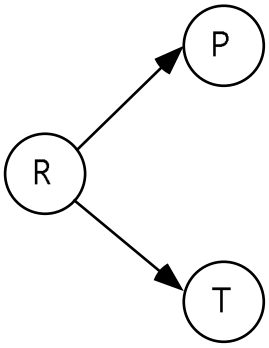

Figure 1: The network showing the relationship between three variables: whether it is raining (R) and whether Peter (P) and Tomas (T) are late.

to get $P(p \mid \bar{r}) = 0.096$.

Similarly for $P(t \mid \bar{r})$:

$$0.05 = (0.5)(0.01) + P(t \mid \bar{r})(0.99)$$

yields $P(t \mid \bar{r}) = 0.045$.

```
(define prof
  (define-bnet 'prof
    '((R () (0.99 0.01))
      (P (R) ((0.904 0.096) (0.5 0.5)))
      (T (R) ((0.955 0.045) (0.5 0.5))))))
```

Part B

If Peter is late, is Tomas more likely to be late?

```
(conditional-probability prof '((T #t)) '((P #t)))
=> .06774090363854457
```

which is higher than his *a priori* probability of being late, 0.05. In fact, observing that Peter is late makes it more likely that it is raining (the only cause we have for his being late), and if it is indeed raining, that makes it more likely that Tomas will be late as well.

The above argument is symmetric, so the probability that Peter is late will depend on whether Tomas is late. Indeed, using our definition of `independent?` from part ,

```
(independent? prof 'P 'T)
=> ()
```

If we know that it is raining, then there is nothing to be learned about either Peter or Tomas being late from learning that the other one is. Therefore, they are conditionally independent. The probability that each is late is like a separate coin toss.

This is a common feature of Bayes nets: two nodes with a single common ancestor will, in general, be dependent, but be made independent by knowing the value of the ancestor. (This generally depends on their not being related along other paths in the network.)

```
(c-independent? prof 'P 'T '((R #t)))
=> #t
```

Part C

```
(define (c-independent? bnet var1 var2 circ)
  (let ((res #t))
    (dolist '(#f #t)                    ;assumes T/F vars!
      (lambda (val1)
        (dolist '(#f #t)
          (lambda (val2)
            (set! res
              (and
               res
               (< (abs (- (* (conditional-probability
                               bnet
                               (list (list (var1 val1)))
                               circ)
                              (conditional-probability
                               bnet
                               (list (list (var2 val2)))
                               circ))
                           (conditional-probability
                            bnet
                            (list (list (var1 val1))
                                  (list (var2 val2)))
                            circ)))
                  0.001))))))
    res))

(c-independent? mcbn1 'A 'B '((C #t)))
=> ()

(c-independent? mcbn1 'D 'E '((C #t)))
=> #t
```

```
(c-independent? mcbn1 'A 'D '((B #f) (C #t)))
=> #t

(c-independent? mcbn1 'A 'D '((B #f) (C #f)))
=> #t
```

The definition of conditional independence is

$$P(A, B \mid C) = P(A \mid C)P(B \mid C).$$

For two variables, A and B, to be independent given knowledge of C, the relationship must hold for all possible values of the variables, so to check for conditional independence, we need to verify that

$$\forall_{x,y,z} P(A = x, B = y \mid C = z) = P(A = x \mid C = z)P(B = y \mid C = z).$$

In the way we have formulated the problem, however, we ask only whether two variables are conditionally independent given a particular circumstance (recall: an assignment of values to some variables). In this (weaker) check, we need not quantify over all the values of the variables assigned in the circumstance, but check only for the specific values that are assigned. This is why `c-independent?` iterates over the values of variables `var1` and `var2`, but not the variables in `circ`. It might have been better to define `(c-independent? bnet var1 var2 vars)` to check that `var1` and `var2` are independent (for all their possible values) given all possible values of the variables in the list `vars`.

Problem 5 Modeling Complex Probabilistic Dependence

Part A Building the Model

The network that describes Holmes' initial problem formulation involves four binary random variables, and thus four nodes in the Bayes network:

B whether Holmes' house has in fact been burlarized (burgled?)

A whether his alarm sounded

W whether Watson telephones

G whether Gibbons telephones

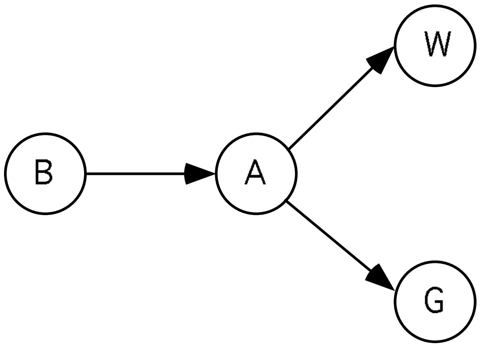

Figure 2: Bayes net for the Holmes burglary problem. See text for legend.

We may reasonably assume that B depends probabilistically only on the intent of some burglar to break in, so we model this simply by giving an *a priori* probability of the burglary, $P(b) = 10^{-4}$. (Recall that $P(b)$ abbreviates $P(B = \texttt{\#t})$.)

It also seems reasonable to model the alarm as depending only on whether the burglary actually took place. If it did, the alarm would go off 99% of the time; if it did not, the alarm would nevertheless go off 0.2% of the time. We may also assume that both Watson's and Gibbon's telephoning depend only on the alarm and not on the burglary itself—i.e., that G and W get their knowledge about the burglary only through whether there was an alarm or not. We also assume that G and W are conditionally independent, i.e., given that the alarm sounds, Gibbons is neither more nor less likely to call depending on whether Watson calls or not, and similarly for the case where the alarm does not sound.

When Gibbons calls Holmes, she is drunk and tells him some long, complicated story that makes him believe more strongly than he might otherwise that the alarm at his house is going off. Presumably, there is some actual probability that Gibbons would tell this particular story if the alarm were ringing, $P(g|a)$, and some other, smaller, probability that she would tell exactly this same story if the alarm were not ringing, $P(g|\bar{a})$. Now it is probably quite difficult to determine just what these probabilities really are, unless you have a very detailed psychological profile of Gibbons, a theory of how she constructs telephone stories while drunk, and how these depend on whether or not an alarm is sounding from Holmes' house. Practically speaking, neither we nor Holmes could accurately estimate these two conditional probabilities.

The observation in the assignment was that despite the difficulty of getting those numbers, Holmes might have a much more intuitively robust sense that Gibbons would be five times as likely to be telling him this story if the alarm were sounding than otherwise. If you recall the discussion of the odds/likelihood ratio form of Bayes Rule, we can adequately represent Holmes' intuition just by the likelihood ratio, $P\frac{(g|a)}{P(g|\bar{a})}$, even if we don't know either particular number. Therefore, we can arbitrarily take, say, $P(g|a)$ to be 0.5 and $P(g|\bar{a})$ to be 0.1, because they have the ratio 5. The fact that we are not making accurate estimates of these conditional probabilities does mean that we could not use the same network to predict how likely Gibbons is to tell this particular story to Holmes under various circumstances.

Once she has done so, however, our model allows us to update our likelihood that the alarm actually sounded.

B=#t	B=#f
0.0001	0.9999

B	A=#t	A=#f
#t	0.99	0.01
#f	0.002	0.998

A	W=#t	W=#f
#t	0.85	0.15
#f	0.25	0.75

A	G=#t	G=#f
#t	0.50	0.50
#f	0.10	0.90

This leads to the following net definition:

```
(define holmes
  (define-bnet 'holmes
    '((B () (0.9999 0.0001))
      (A (B) ((0.998 0.002) (0.01 0.99)))
      (W (A) ((0.75 0.25) (0.15 0.85)))
      (G (A) ((0.90 0.10) (0.5 0.5))))))
```

Part B Calculations

```
(conditional-probability holmes '((B #t)) '((W #t) (G #t)))
=> 1.6292872313417586e-3
```

so getting the call from Watson and Gibbons raises Holmes' estimate of the probability of burglary from 0.0001 to 0.0016, a factor of 16.

Part C Explaining Away

We add nodes to the network above to represent the presence or absence of the earthquake (E) and the radio report (K). In this network, we have a node (A) that depends on two parent nodes, not just one. The assignment suggests that you model the relationship $P(A \mid B, E)$

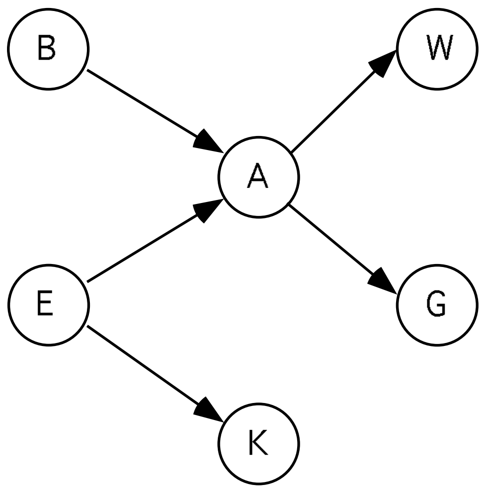

Figure 3: Bayes net for the Holmes burglary problem, with the addition of nodes for earthquake (E) and radio report of quake (K).

as a noisy or, with parameters c_b, c_e and l. I.e., either a burglary or a quake can set off the alarm, but these are independent causes, with a small leak.

The problem suggests that the original model relating burglaries to alarms would serve to let you estimate the leak probability for an alarm happening in the absence of a burglary, namely 0.002. This is $P(A = a|B = \bar{b})$ from the original model. At this point the careful student should say, “Hey there was no mention of Earthquakes in that statement. Shouldn't I assume that some of those alarms were caused by Earthquakes?” Well yes, but we will assume for the moment that you were given direct information about $P(a|\bar{b}\bar{e}) = P_{NO}(a|\bar{b}\bar{e}) = l = 0.002$. We will return to this assumption below.

The problem further states the causal probability of Earthquakes causing Alarms, $c_e = 0.7$. The only remaining parameter is c_b, and finding it will require some algebra. In order to find it we will assume that the probability of an alarm given a burglary, $P(a|b)$, was in fact the probability of an alarm given a burglary and *no* earthquake, $P(a|b\bar{e})$. As a result:

$$
\begin{aligned}
P(a|b) = P(a|b\bar{e}) = P_{NO}(a|b\bar{e}) &= 1 - (1 - c_b)(1 - l) && (1)\\
.99 &= 1 - (1 - c_b)(1 - 0.002) && (2)\\
0.01 &= (1 - c_b) \times .998 && (3)\\
\frac{0.01}{0.998} = 1 - c_b & && (4)\\
c_b = 0.98998 & && (5)
\end{aligned}
$$

You might be tempted to argue that we were running fast and loose with probabilities here. Twice we assumed that a conditional distribution gave us information that another cause was not present. This is almost always a bad idea. It is almost always better to sum over the missing variables as we did earlier.

Let's repeat the argument in a more correct fashion. We are still given c_e the direct causal probability. But now to find both l and c_b we will need to do additional calculations:

$$
\begin{aligned}
1 - 0.002 &= P_{NO}(\bar{a}|\bar{b}) = P_{NO}(\bar{a}|\bar{b}e)P(e) + P_{NO}(\bar{a}|\bar{b}\bar{e})P(\bar{e}) && (6)\\
0.998 &= (1 - c_e)(1 - l) \times 0.0001 + (1 - l) \times 0.9999 && (7)\\
0.998 &= 0.00003(1 - l) + 0.9999(1 - l) && (8)\\
0.998 &= 0.99993(1 - l) && (9)\\
l &= 0.00193 && (10)
\end{aligned}
$$

We also have the following constraint:

$$
\begin{aligned}
1 - 0.99 &= P_{NO}(\bar{a}|b) = P_{NO}(\bar{a}|be)P(e) + P_{NO}(\bar{a}|b\bar{e})P(\bar{e}) && (11)\\
0.01 &= (1 - c_b)(1 - c_e)(1 - l) \times 0.0001 + (1 - c_b)(1 - l) \times 0.9999 && (12)\\
0.01 &= .0000299(1 - c_b) + 0.9979702(1 - c_b) && (13)\\
0.01 &= 0.998(1 - c_b) && (14)\\
0.01002 &= (1 - c_b) && (15)\\
c_b = 0.98998 & && (16)
\end{aligned}
$$

So our first answer was very, very close to the correct answer.

Thus, finally we can specify our model to show that the conditional probability of alarms depends on burglaries and earthquakes by

```
(noisy-or 0.002 0.98998 0.7)
```

The overall model that includes the possible effects of the earthquake and news reports is:

```
(define quake
  (define-bnet 'quake
    '((B () (0.9999 0.0001))
      (E () (0.9999 0.0001))
      (A (B E) (noisy-or 0.002 0.98998 0.7))
      (W (A) ((0.75 0.25) (0.15 0.85)))
      (G (A) ((0.90 0.10) (0.5 0.5)))
      (K (E) ((1.0 0.0) (0.03 0.97))))))
```

The following conditional probabilities show the likelihood of a burglary under various conditions.

```
(conditional-probability quake '((B #t)) '((W #t) (G #t)))
=> 1.6275284697285708e-3

(conditional-probability quake '((B #t)) '((W #t) (G #t) (E #f)))
=> 1.6292872931615268e-3

(conditional-probability quake '((B #t)) '((W #t) (G #t) (K #t)))
=> 1.3883617624894435e-4
```

The first closely matches the result in the model above because earthquakes are generally very unlikely. The second shows that knowing the quake actually did not occur makes little difference, for the same reason. The third gives the desired answer: knowing (from the radio) that the quake occurred, the probability of burglary declines by an order of magnitude; the quake can explain away the alarm without needing to hypothesize the burglary. Note, however, that a little doubt remains and the probability of a burglary is still somewhat higher than its prior probability in the original model.