CS 4100: Foundations of Artificial Intelligence (Spring 2020) Roi Yehoshua
Student name: Zhiyuan Cao (Due) March 19, 2020

Reasoning under Uncertainty

1 Bayesian Networks: The Green Party

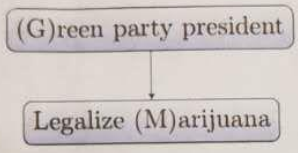
It's election season, and the chosen president may or may not be the Green Party candidate. Pundits believe that Green Party presidents are more likely to legalize marijuana than candidates from other parties, but legalization could occur under any administration. Armed with the power of probability, the analysts model the situation with the Bayes Net below.

Cond. $P(A|B) = \frac{P[A \cap B]}{P(B)}$

Add. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Mul. $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

Ind. $P(A \cap B) = P(A)P(B)$



	+g	-g
P(g)	0.25	0.75

	$P(+m\|g)$	$P(-m\|g)$
+g	0.8	0.2
-g	0.1	0.9

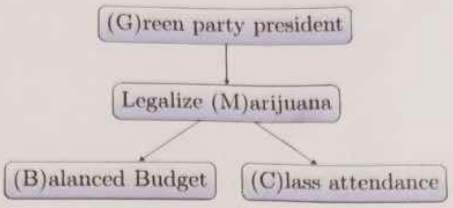
1. What is the marginal probability that marijuana is legalized $P(+m)$?

$$P(+m) = 0.25 \times 0.8 + 0.75 \times 0.1$$
$$= 0.2 + 0.075$$
$$= 0.275$$

2. 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?

$$P(+g|+m) = \frac{P(+g \cap +m)}{P(+m)} = \frac{P(+m|+g)P(g)}{0.275} = \frac{0.8 \cdot 0.25}{0.275} = \frac{0.2}{0.275}$$
$$= \frac{20}{275}$$
$$= \frac{8}{11}$$

We can make better inferences if we observe more evidence. Now we are going to 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.



	$P(+b\|m)$	$P(-b\|m)$
+m	0.3	0.7
-m	0.2	0.8

	$P(+c\|m)$	$P(-c\|m)$
+m	0.2	0.8
-m	0.5	0.5

3. Complete the full joint distribution table given below:

G	M	B	C	$P(G,M,B,C)$
+	+	+	+	0.012
+	+	+	-	0.048
+	+	-	+	0.028
+	+	-	-	0.112
+	-	+	+	0.005
+	-	+	-	0.005
+	-	-	+	0.07
+	-	-	-	0.02
-	+	+	+	0.0045
-	+	+	-	0.018
-	+	-	+	0.0105
-	+	-	-	0.042
-	-	+	+	0.0675
-	-	+	-	0.0675
-	-	-	+	0.27
-	-	-	-	0.27

$0.25 \cdot 0.8 \cdot 0.3 \cdot 0.7$
0.8

$0.25 \cdot 0.2 \cdot 0.2 \cdot 0.5$

$0.25 \cdot 0.2 \cdot 0.8$

$0.75 \cdot 0.1 \cdot 0.3 \cdot 0.7$

$0.75 \cdot 0.1 \cdot 0.7 \cdot 0.8$

$0.75 \cdot 0.9 \cdot 0.8 \cdot 0.5$

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

(a) $P(+b|+m)$ = 0.3

(b) $P(+b|+m,+g)$ = P(+b|+m) = 0.3 $b \perp M$

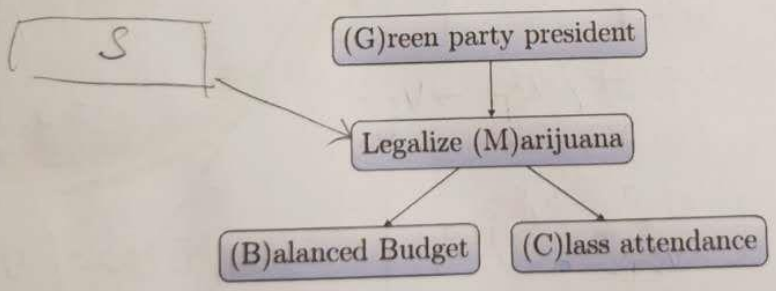
(c) $P(+b)$ = 0.012 + 0.048 + 0.05 + 0.25 + 0.0045 + 0.018 + 0.0675 + 0.0675

= 0.06 + 0.01 + 0.0225 + 0.135 = 0.07 + 0.1575 = 0.2275

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

$= \frac{P(+c \wedge +b)}{P(+b)} = \frac{0.012 + 0.05 + 0.0045 + 0.0675}{0.2275} = \frac{0.088}{0.2275} = \frac{88}{227.5} = \frac{176}{455}$

5. Now, add a node S to the Bayes net 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. Draw the new Bayes net below. Which CPT or CPTs need to be modified?



P(M|G) → P(M|G,S)

And there are more entries to correct S negative and positive

6. Consider your augmented model. Just based on the structure, which of the following are guaranteed to be true, and which are not guaranteed to be true?

(a) $B \perp\!\!\!\perp G$

(b) $B \perp\!\!\!\perp C$

(c) $C \perp\!\!\!\perp G|M$ ✓ After observe M

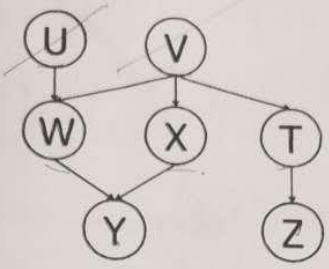
(d) $B \perp\!\!\!\perp C|G$

(e) $G \perp\!\!\!\perp S$ ✓

(f) $G \perp\!\!\!\perp S|B$

2 Variable Elimination

For the Bayes net below, we are given the query $P(Z|+y)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: U, V, W, T, X.



Complete the following description of the factors generated in this process: After inserting evidence, we have the following factors to start out with:

$$P(U), P(V), P(W|U,V), P(X|V), P(T|V), P(+y|W,X), P(Z|T)$$

When eliminating U we generate a new factor f_1 as follows:

$$f_1(V,W) = \sum_u P(u)P(W|u,V)$$

This leaves us with the factors:

$$P(V), P(X|V), P(T|V), P(+y|W,X), P(Z|T), f_1(V,W)$$

1. When eliminating V we generate a new factor f_2 as follows:

$$f_2(W,X,T) = \sum_v P(v)P(X|v)P(T|v)f_1(v,W)$$

2. This leaves us with the factors:

$$f_2(W,X,T) \quad P(+y|W,X), \quad P(Z|T)$$

3. When eliminating W we generate a new factor f_3 as follows:

$$f_3(X, T, Y) = \sum_W f_2(W, X, T) \cdot P(Y \mid W, X)$$

4. This leaves us with the factors:

$$f_3(X, T, Y), \quad P(Z \mid T)$$

5. When eliminating T we generate a new factor f_4 as follows:

$$f_4(X, Y, Z) = \sum_T f_3(X, T, Y) \cdot P(Z \mid T)$$

6. This leaves us with the factors:

$$f_4(X, Y, Z)$$

7. When eliminating X we generate a new factor f_5 as follows:

$$f_5(Y, Z) = \sum_X f_4(X, Y, Z)$$

8. This leaves us with the factors:

$$f_5(Y, Z)$$

9. Briefly explain how $P(Z|+y)$ can be computed from f_5.

we just need to renormalize $f_5(Z,+y)$

$$P(Z|+y) = \frac{f_5(Z,+y)}{\sum_z f_5(z,+y)}$$

10. Among $f_1, f_2, \cdots, f_5$ which is the largest factor generated? (Assume all variables have binary domains.) How large is this factor?

f_2 has the largest one

$2^3 = 8$

11. Find a variable elimination ordering for the same query, i.e., for $P(Z|y)$, for which the maximum size factor generated along the way is smallest. Hint: the maximum size factor generated in your solution should have only 2 variables, for a size of $2^2 = 4$ table. Fill in the variable elimination ordering and the factors generated into the table below

Note: in the naive ordering we used earlier, the first line in this table would have had the following two entries: $U, f_1(V, W)$.

Variable Eliminated	Factor Generated
U	$f_1(V,W)$
T	$f_2(V,Z)$
W	$f_3(V,+y,X)$
X	$f_4(V,Y)$
V	$f_5(Z,Y)$

3 HMM

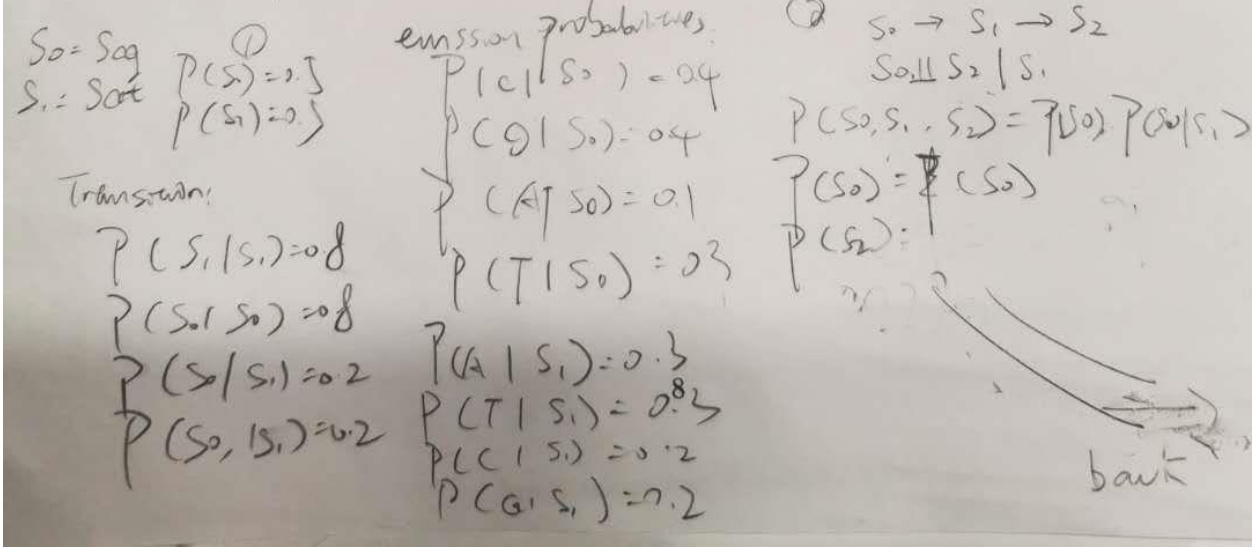
A DNA sequence is a series of components from $\{A, C, G, T\}$. Suppose there is one hidden variable S that controls the generation of DNA sequence. S can have 2 possible states: $\{cg, at\}$. Both states are equally likely initially. At each time step, the probability of transitioning to the other state is 0.2; otherwise, the state remains the same. In state cg, the probability of observing each of C or G is 0.4, and the probability of observing each of A and T is 0.1. In state at, the probability of observing each of A or T is 0.3, and the probability of observing each of C and G is 0.2.

1. Give a precise formulation of this HMM (initial belief, transition and observation probabilities).

2. Recall that an HMM assumes the Markov assumption $S_t \perp\!\!\!\perp S_{t-k} \mid S_{t-1}$ for $k > 1$. Hence we know that $S_2 \perp\!\!\!\perp S_0 \mid S_1$. Show that this *does not* imply that S_2 is independent of S_0, i.e., show that $S_2 \not\perp\!\!\!\perp S_0$. This therefore shows that conditional independence does not imply independence.
   Hint: Using the provided model, write out the joint distribution $P(S_0, S_1, S_2)$. Then compute the marginal distributions $P(S_0)$, $P(S_2)$, and $P(S_0, S_2)$, and show that these marginal distributions do not satisfy the definition of independence.

3. Suppose we observed the sequence $\mathbf{e} = TGCACA$. Using the HMM filtering equations, compute $\mathbf{P}(S_6|\mathbf{E} = \mathbf{e})$. Show your work.

Note: This is not just an arbitrary problem in HMMs. The model above is a simplified version of a model for detecting CpG islands (regions with high frequency of CG nucleotides) in the genome, which is a classic application of HMMs in bioinformatics. CpG islands are biologically important because in nature these CG nucleotides tend to change into CA pairs over evolutionary time via methylation, which means that we should expect CpG islands to be rare. The fact that CpG islands exist in certain regions of the genome therefore indicate that those regions are for some reason *important* for survival, because these regions must have been preserved by natural selection. (Consider the mutants whose CG nucleotides in this region have been converted to CA: they must have all died off, which is why we only observe the CG variants).



② $S_0 \rightarrow S_1 \rightarrow S_2$

From Markov rule, we know $S_2 \perp S_0 \mid S_1$

$P(S_0, S_1, S_2) = P(S_0)\, P(S_1 \mid S_0)\, P(S_2 \mid S_1)$

if $P(S_0)$ and $P(S_1)$ are independent,

$P(S_0, S_2) = P(S_0) \cdot P(S_2)$

$P(S_0 = at) = 0.5 \quad P(S_0 = cg) = 0.5$

suppose $S_0 = cg \quad P(S_0) = 0.5$

$S_1 = at \quad P(S_1) = 0.5 \cdot 0.2 = 0.1$

$S_2 = at \quad P(S_2) = 0.08$

S_0 \ S_1	at	cg
at	0.5	0.5
cg	0.5	0.5

$P(S_0) \cdot P(S_2) = 0.5 \cdot 0.08$

$= 0.04$

But $P(S_0, S_2) = 1 \neq 0.04$

so, it's not imply independent if S_1 is not given

3. When eliminat...

$f_{3?}(x,T,+) = \sum_w f_2(w,x,T)$ ...

TGCACA

③ Initial $\forall st \in \{Scg, Sct\}$

S_{cg}	S_{ct}
$P(S_{cg})$ 0.5	$P(S_{ct})$ 0.5
$P(S_{cg}\mid e=T)$ $0.05/(0.05+0.15)=0.25$	$P(S_{ct}\mid e=T)$ $0.15/(0.05+0.15)=0.75$
$P(S_{cg}\mid e=TG)$ $(0.25\cdot 0.8+0.75\cdot 0.2)\cdot 0.4$ $=0.14$ $0.14/(0.14+0.13)=0.5185$	$P(S_{ct}\mid e=TG)\propto (0.25\cdot 0.2+0.75\cdot 0.8)\cdot 0.2=0.13$ $0.13/(0.14+0.13)=0.48148$
$P(S_{cg}\mid e=TGC)$ $(0.5185\cdot 0.8+0.48148\cdot 0.2)\cdot 0.4$ $=0.2044$ $0.2044/(0.2044+0.0978)$ $=0.676$	$P(S_{ct}\mid e=TGC)$ $(0.5185\cdot 0.2+0.48148\cdot 0.8)\cdot 0.2$ $=0.0978$ $0.0978/(0.2044+\cdots)$ $=0.324$
$P(S_{cg}\mid e=TGCA)$ $(0.676\cdot 0.8+0.324\times 0.2)\cdot 0.1$ $=0.067$ $0.067/(0.067+0.118)$ $=0.34$	$P(S_{ct}\mid e=TGCA)$ $(0.676\cdot 0.2+0.324\cdot 0.8)\cdot 0.3$ $=0.118$ $0.118/(0.067+0.118)=0.66$
$P(S_{cg}\mid e=TGCAC)$ $(0.34\cdot 0.8+0.66\cdot 0.2)\cdot 0.4$ $=0.1616$ $0.1616/(0.1616+0.1192)$ $=0.575$	$P(S_{ct}\mid e=TGCAC)$ $(0.34\cdot 0.2+0.66\cdot 0.8)\cdot 0.2$ $=0.1192$ $0.1192/(0.1616+0.1192)$ $=0.425$
$P(S_{cg}\mid e=TGCACA)$ $(0.575\cdot 0.8+0.425\cdot 0.2)\cdot 0.1$ $=0.0545$ $0.0545/(0.0545+0.1365)$ $=0.285$	$P(S_{ct}\mid e=TGCACA)$ $(0.575\cdot 0.2+0.425\cdot 0.8)\cdot 0.3$ $=0.1365$ $0.1365/(0.0545+0.1365)$ $=0.715$