Q1.	3-Player Games	/8
Q2.	HMMs and Particle Filtering	/12
Q3.	Occupy Cal	/12
Q4.	Instantiated Elimination	/14
Q5.	Bayes' Net Representation	/12
Q6.	Exponential Utilities	/14
Q7.	Bayes Net CSPs	/9
Q8.	Q-Learning Strikes Back	/8
Q9.	Probability and Bayes Nets	/11
	Total	/100

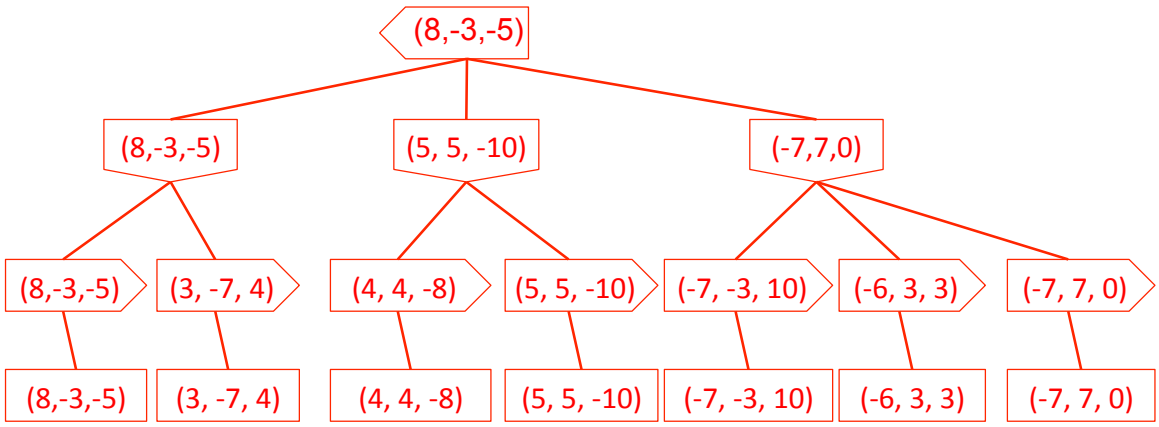
DATA 130008 Introduction to
Spring 2017 Artificial Intelligence

Quiz Sample

Q1. [8 pts] 3-Player Games

(a) [4 pts] **A 3-Player Game.**

Consider the 3-player game shown below. The player going first (top of the tree) is the Left player, the player going second is the Middle player, and the player going last is the Right player, optimizing the left, middle and right utility value respectively. Fill in the values at all nodes. Note all players *maximize* their respective utility value shown in the tree.



(b) [4 pts] **Pruning for Zero-Sum 3-Player Game.** The same game tree is shown again below.

Now assume that we have the knowledge that the sum of the utilities of all 3 players is always zero. Under this assumption is any pruning possible similar to $\alpha - \beta$ pruning? If so mark the pruning on the tree above. If not, briefly explain why not below.

No. $\alpha - \beta$ pruning works based on the following mechanism: when considering options for the minimizer at a node n, you can prune if you found an option for the minimizer that is worse than something the maximizer could ensure by diverting higher-up in the tree such that n would never be reached. Reason: no matter what the other options are at node n, diverting will be better for the maximizer.

In the three player game we cannot perform such inference: if, let's say, the right player while considering options at a node n has an option that is not great for the left player, let's say giving the left player a pay-off of x, then it is still possible for the right player to prefer another option at node n which is better than x for the left-player.

Q2. [12 pts] HMMs and Particle Filtering

Consider a Markov Model with a binary state X (i.e., X_t is either 0 or 1). The transition probabilities are given as follows:

X_t	X_{t+1}	$P(X_{t+1} \mid X_t)$
0	0	0.9
0	1	0.1
1	0	0.5
1	1	0.5

(a) [2 pts] The prior belief distribution over the initial state X_0 is uniform, i.e., $P(X_0 = 0) = P(X_0 = 1) = 0.5$. After one timestep, what is the new belief distribution, $P(X_1)$?

X_1	$P(X_1)$
0	0.7
1	0.3

Since the prior of X_0 is uniform, the belief at the next step is the transition distribution from X_0:
$p(X_1 = 0) = p(X_0 = 0)p(X_1 = 0|X_0 = 0) + p(X_0 = 1)p(X_1 = 0|X_0 = 1) = .5(.9) + .5(.5) = .7.$
$p(X_1 = 1) = p(X_0 = 0)p(X_1 = 1|X_0 = 0) + p(X_0 = 1)p(X_1 = 1|X_0 = 1) = .5(.1) + .5(.5) = .3.$

(b) [2 pts] Now, we incorporate sensor readings. The sensor model is parameterized by a number $\beta \in [0, 1]$:

X_t	E_t	$P(E_t \mid X_t)$
0	0	β
0	1	$(1 - \beta)$
1	0	$(1 - \beta)$
1	1	β

(c) [2 pts] At $t = 1$, we get the first sensor reading, $E_1 = 0$. Use your answer from part (a) to compute $P(X_1 = 0 \mid E_1 = 0)$. Leave your answer in terms of β.

$$p(X_1 = 0|E_1 = 0) = \frac{p(E_1 = 0|X_1 = 0)p(X_1 = 0)}{\sum_x p(E_1 = 0|X_1 = x)p(X_1 = x)}$$
$$= \frac{\beta(0.7)}{\beta(0.7) + (1 - \beta)(0.3)}$$

(d) [2 pts] For what range of values of β will a sensor reading $E_1 = 0$ increase our belief that $X_1 = 0$? That is, what is the range of β for which $P(X_1 = 0 \mid E_1 = 0) > P(X_1 = 0)$?
$\beta \in (0.5, 1]$. Intuitively, observing $E_1 = 0$ will only increase the belief that $X_1 = 0$ if $E_1 = 0$ is more likely under $X_1 = 0$ than not. Note that $\beta > 0.5$; $\beta = 0.5$ is uninformative since the conditional distribution is uniform. This can be verified algebraically by setting $p(X_1 = 0) = p(X_1 = 0|E_1 = 0)$ and solving for β.

(e) [2 pts] Unfortunately, the sensor breaks after just one reading, and we receive no further sensor information. Compute $P(X_\infty \mid E_1 = 0)$, the stationary distribution *very many* timesteps from now.

X_∞	$P(X_\infty \mid E_1 = 0)$
0	$\frac{5}{6}$
1	$\frac{1}{6}$

The stationary distribution π for transition matrix P satisfies $\pi = \pi P$. It is purely a function of the transition matrix and not the prior or past observations.

Determine π by setting up the matrix equation $\pi = \pi P$ with the additional equation $\pi_0 + \pi_1 = 1$ from the sum-to-one constraint of the probabilities and solving.

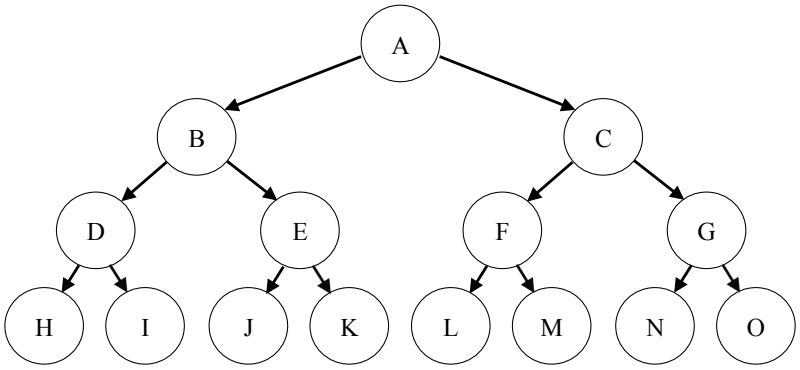
(f) [2 pts] How would your answer to part (d) change if we never received the sensor reading E_1, i.e. what is $P(X_\infty)$ given no sensor information?

X_∞	$P(X_\infty)$
0	$\frac{5}{6}$
1	$\frac{1}{6}$

The stationary distribution does not depend on past observations, so the distribution is unchanged whether E_1 is observed or not.

Q3. [12 pts] Occupy Cal

(a) [3 pts] You are at Occupy Cal, and the leaders of the protest are deciding whether or not to march on California Hall. The decision is made centrally and communicated to the occupiers via the "human microphone"; that is, those who hear the information repeat it so that it propagates outward from the center. This scenario is modeled by the following Bayes net:



A	*P(A)*
+m	0.5
-m	0.5

π(X)	*X*	*P(X|π(X))*
+m	*+m*	0.9
+m	*-m*	0.1
-m	*+m*	0.1
-m	*-m*	0.9

Each random variable represents whether a given group of protestors hears instructions to march ($+m$) or not ($-m$). The decision is made at A, and both outcomes are equally likely. The protestors at each node relay what they hear to their two child nodes, but due to the noise, there is some chance that the information will be misheard. Each node except A takes the same value as its parent with probability 0.9, and the opposite value with probability 0.1, as in the conditional probability tables shown.

(b) [3 pts] Compute the probability that node A sent the order to march ($A = +m$) given that both B and C receive the order to march ($B = +m$, $C = +m$).

$$
\begin{aligned}
p(A = +m \mid B = +m, C = +m) &= \frac{p(A = +m, B = +m, C = +m)}{\sum_a p(A = a, B = +m, C = +m)} \\
&= \frac{p(A = +m)p(B = +m|A = +m)p(C = +m|A = +m)}{\sum_a p(A = a)p(B = +m|A = a)p(C = +m|A = a)} \\
&= \frac{p(A = +m)p(X = +m|\pi(X) = +m)^2}{\sum_a p(A = a)p(X = +m|\pi(X) = a)^2} \\
&= \frac{.5 \cdot .9^2}{.5 \cdot .9^2 + .5 \cdot .1^2)} = \frac{.9^2}{.9^2 + .1^2} \approx 0.988
\end{aligned}
$$

.

Note that $P(A)$ is uniform. It can be pulled out of sums and cancelled in the numerator and denominator. For simplicity $P(A)$ terms are dropped.

For further simplification, note that the conditional distribution $P(X|\pi(X))$ is symmetric, so calculations can be simplified in terms of $p(X, \pi(X)same) = .9$ and $(p(X, \pi(X)different) = .1$.

(c) [3 pts] Compute the probability that D receives the order $+m$ given that A sent the order $+m$.

$$\begin{aligned} p(D = +m|A = +m) &= \frac{\sum_b p(D = +m, B = b, A = +m)}{\sum_d \sum_b p(D = +m, B = b, A = +m)} \\ &= \frac{\sum_b p(D = +m|B = b)p(B = b|A = +m)}{\sum_d \sum_b p(D = d|B = b)p(B = b|A = +m)} \\ &= \frac{.9^2 + .1^2}{.9^2 + .1^2 + 2(.9)(.1)} = 0.82 \end{aligned}$$

You are at node D, and you know what orders have been heard at node D. Given your orders, you may either decide to march (*march*) or stay put (*stay*). (Note that these actions are distinct from the orders $+m$ or $-m$ that you hear and pass on. The variables in the Bayes net and their conditional distributions still behave exactly as above.) If you decide to take the action corresponding to the decision that was actually made at A (not necessarily corresponding to your orders!), you receive a reward of $+1$, but if you take the opposite action, you receive a reward of -1.

(d) [3 pts] Given that you have received the order $+m$, what is the expected utility of your optimal action? (Hint: your answer to part (b) may come in handy.)

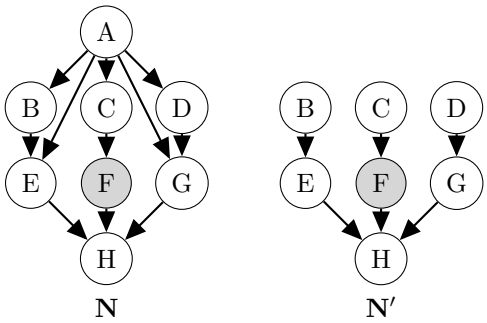
The maximum expected utility is 0.64. The expected utility of an action is $p(A = action|D = +m)(1) + p(A \neq action|D = +m)(-1)$, or the expected reward of matching plus the expected reward of not matching.

The conditional probability $p(D = +m|A = +m)$, the answer to part (b), is the probability that the order at D matches the order given at A. This follows from the symmetry of the distribution of A and the conditionals of the children (this can be verified by computing $p(A = +m|D = +m)$).

For marching, the expected utility is $p(D = +m|A = +m)(1) + p(D = -m|A = +m)(-1) = .82(1) + (1 - .82)(-1) = 0.64$. For standing, the expected utility is $p(D = +m|A = +m)(-1) + p(D = -m|A = +m)(1) = .82(-1) + (1 - .82)(1) = -0.64$.

Q4. [14 pts] Instantiated Elimination

(a) Difficulty of Elimination. Consider answering $P(H \mid +f)$ by variable elimination in the Bayes' nets N and N'.
Elimination order is alphabetical.
All variables are binary $+/-$.
Factor size is the number of unobserved variables in a factor made during elimination.



(i) [2 pts] What is the size of the largest factor made during variable elimination for N?

5

$F(B, C, D, E, G)$ after eliminating A. 2^5, the number of factor entries was also accepted.

(ii) [2 pts] What is the size of the largest factor made during variable elimination for N'?

2

$F(G, H, +f)$ after eliminating E.

Variable elimination in N can take a lot of work! If only A were observed...

(b) Instantiation Sets. To simplify variable elimination in N, let's pick an *instantiation set* to pretend to observe, and then do variable elimination with these additional instantiations.

Consider the original query $P(H \mid +f)$, but let A be the instantiation set so $A = a$ is observed. Now the query is H with observations $F = +f, A = a$.

(i) [2 pts] What is the size of the largest factor made during variable elimination with the $A = a$ instantiation? 2

Observing $A = a$ renders the children conditionally independent and prevents its elimination. The large factor $F(B, C, D, E, G)$ is never made and elimination proceeds similarly to elimination in N' except for the presence of a in the factors.

(ii) [1 pt] Given a Bayes' net over n binary variables with k variables chosen for the instantiation set, how many instantiations of the set are there?

2^k

For a selected instantiation set of k binary variables, there are 2^k total instantiations of these variables. The alternative interpretation of how many size k instantiation sets exist from n variables, $\binom{n}{k}$ was also accepted.

(c) Inference by Instantiation. Let's answer $P(H| + f)$ by variable elimination with the instantiations of A.

(i) [2 pts] What quantity does variable elimination for $P(H| + f)$ with the $A = +a$ instantiation compute *without normalization*? That is, which choices are equal to the entries of the last factor made by elimination?

- ○ $P(H \mid +f)$
- ● $P(H, +a, +f)$
- ○ $P(H, +f \mid +a)$
- ○ $P(H \mid +a)$
- ○ $P(H, +a \mid +f)$
- ○ $P(H \mid +a, +f)$

At the end of variable elimination, the last factor is equal to the equivalent entries of the joint distribution with the eliminated variables summed out and the selected values of the evidence variables. Normalization gives the conditional $p(+h \mid +a, +f) = \frac{f(+h,+a,+f)}{f(+h,+a,+f)+f(-h,+a,+f)}$, but here the factor is kept unnormalized.

(ii) [2 pts] Let $I_+(H) = F(H, +a, +f)$ and $I_-(H) = F(H, -a, +f)$ be the last factors made by variable elimination with instantiations $A = +a$ and $A = -a$. Which choices are equal to $p(+h \mid +f)$?

- ○ $I_+(+h) \cdot p(+a) \cdot I_-(+h) \cdot p(-a)$
- ○ $I_+(+h) \cdot p(+a) + I_-(+h) \cdot p(-a)$
- ○ $I_+(+h) + I_-(+h)$
- ○ $\frac{I_+(+h) \cdot p(+a) \cdot I_-(+h) \cdot p(-a)}{\sum_h I_+(h) \cdot p(+a) \cdot I_-(h) \cdot p(-a)}$
- ○ $\frac{I_+(+h) \cdot p(+a) + I_-(+h) \cdot p(-a)}{\sum_h I_+(h) \cdot p(+a) + I_-(h) \cdot p(-a)}$
- ● $\frac{I_+(+h) + I_-(+h)}{\sum_h I_+(h) + I_-(h)}$

The last factors are the entries from the corresponding joint $I_+(+h) = p(+h, +a, +f)$ and $I_-(+h) = p(+h, -a, +f)$ and so on.

By the law of total probability $p(+h, +f) = p(+h, +a, +f) + p(+h, -a, +f)$, so the joint of the original query and evidence can be computed from the instantiated elimination factors. For the conditional $p(+h, +f)$, normalize by the sum over the query $\sum_h p(h, +f) = \sum_h \sum_a p(h, a, +f) = f(+h, +a, +f) + f(+h, -a, +f) + f(-h, +a, +f) + f(-h, -a, +f)$ where the joint over the query and evidence is again computed from the law of total probability over A.

Working with the joint in this way is why the last factor of instantiated elimination was left unnormalized. (Those who answered the conditional $P(H \mid +a, +f)$ in the previous part were awarded credit for multiplying the marginal $p(+a), p(-a)$ back in as this is the correct chain rule were $P(H \mid +a, +f)$ correct.)

(d) [3 pts] **Complexity of Instantiation.** What is the time complexity of instantiated elimination? Let n = number of variables, k = instantiation set size, f = size of the largest factor made by elimination without instantiation, and i = size of the largest factor made by elimination with instantiation. Mark the tightest bound. Variable elimination without instantiation is $O(n \exp(f))$.

- ○ $O(n \exp(k))$
- ○ $O(n \exp(i))$
- ● $O(n \exp(i + k))$
- ○ $O(n \exp(f))$
- ○ $O(n \exp(f - k))$
- ○ $O(n \exp(i/f))$

To carry out instantiated elimination, we have to do variable elimination $\exp(k)$ times for all the settings of the instantiation set. Each of these eliminations takes time bounded by $n \exp(i)$ as the largest factor is the most expensive to eliminate and there are n variables to eliminate.

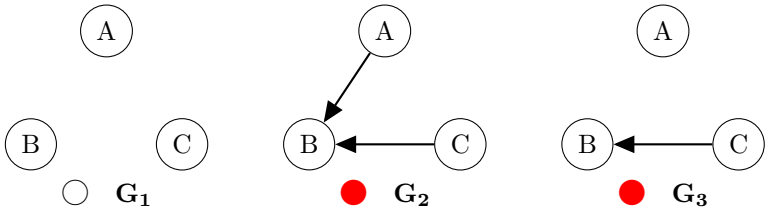
If the instantiation set is not too large and the size of the factors made by instantiation elimination are small enough this method can be exponentially faster than regular elimination. The catch is how to select the instantiation set.

Q5. [12 pts] Bayes' Net Representation

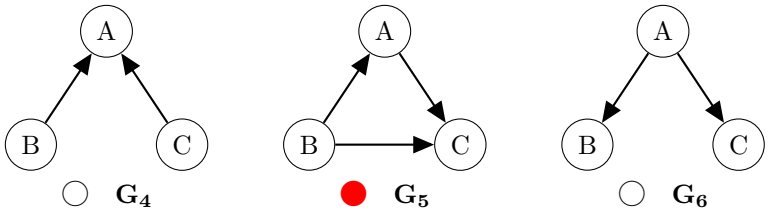
(a) [4 pts] Consider the joint probability table on the right.

Clearly fill in all circles corresponding to BNs that can correctly represent the distribution on the right. If no such BNs are given, clearly select *None of the above.*

A	B	C	P(A, B, C)
0	0	0	.15
0	0	1	.1
0	1	0	0
0	1	1	.25
1	0	0	.15
1	0	1	.1
1	1	0	0
1	1	1	.25



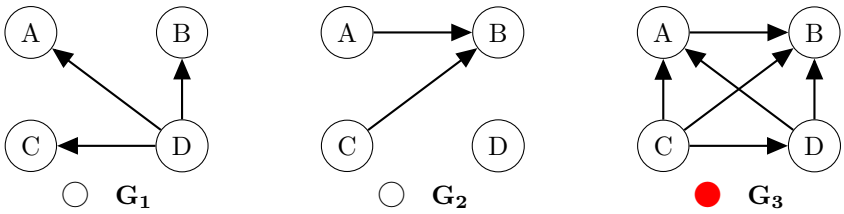
○ **G1** ● **G2** ● **G3**



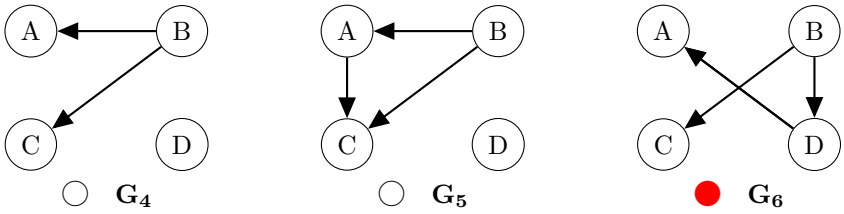
○ **G4** ● **G5** ○ **G6** ○ **None of the above.**

From the table we can clearly see that the values of $P(A, B, C)$ repeat in two blocks. This means that the value of A does not matter to the distribution. The values are otherwise all unique, meaning that there is a relationship between B and C. Together, this means that $A \perp\!\!\!\perp B$, $A \perp\!\!\!\perp B \mid C$, $A \perp\!\!\!\perp C$, and $A \perp\!\!\!\perp C \mid B$ are the only independences in the distribution.

(b) [4 pts] You are working with a distribution over A, B, C, D that can be fully represented by just three probability tables: $P(A \mid D)$, $P(C \mid B)$, and $P(B, D)$. Clearly fill in the circles of those BNs that can correctly represent this distribution. If no such BNs are given, clearly select *None of the above.*



○ **G1** ○ **G2** ● **G3**



○ **G4** ○ **G5** ● **G6** ○ **None of the above.**

The only independences guaranteed by the factorization $P(A, B, C, D) = P(A \mid D)P(C \mid B)P(B, D)$ are:

- $A \perp\!\!\!\perp B \mid D$
- $A \perp\!\!\!\perp C \mid B$
- $A \perp\!\!\!\perp C \mid D$
- $C \perp\!\!\!\perp D \mid B$

(c) [4 pts] We are dealing with two probability distributions over N variables, where each variable can take on exactly d values. The distributions are represented by the two Bayes' Nets shown below. If S is the amount of

storage required for the CPTs for $X_2, \ldots, X_N$ in D_1, how much storage is required for the CPTs for $X_2, \ldots, X_N$ in D_2? **There is a correct answer among the options.**

D_1 **D_2**

- ○ S
- ○ 2^S
- ● Sd
- ○ S^2
- ○ S^d
- ○ $S + 2^d$

D_1 needs storage of Nd values. D_2 needs storage of Nd^2 values.

Q6. [14 pts] Exponential Utilities

(a) The ghosts offer Pacman a deal: upon rolling a fair 6-sided die, they will give Pacman a reward equal to the number shown on the die minus a fee x, so he could win $1-x, 2-x, 3-x, 4-x, 5-x$ or $6-x$ with equal probability. Pacman can also refuse to play the game, getting 0 as a reward.

(i) [2 pts] Assume Pacman's utility is $U(r) = r$. Pacman should accept to play the game if and only if:

○ $x \leq 7/6$ ● $x \leq 7/2$ ○ $x \leq 21/2$ ○ $x \leq 21$

We consider when $EU(play) \geq EU(stay)$.
$EU(stay) = 0$.
$EU(play) = \frac{1}{6}[(1-x) + (2-x) + (3-x) + (4-x) + (5-x) + (6-x)]$
$\frac{1}{6}(21 - 6x) \geq 0$ when $x \leq \frac{7}{2}$. Answer: $x \leq \frac{7}{2}$.

(ii) [2 pts] Assume Pacman's utility is $U'(r) = 2^r$. Pacman should accept to play the game if and only if:

○ $x \leq \log_2(7/2)$ ○ $x \leq \log_2(20)$ ● $x \leq \log_2(21)$ ○ $x \leq 21$

We consider when $EU(play) \geq EU(stay)$.
$EU(stay) = 2^0 = 1$.
$EU(play) = \frac{1}{6}[2^{1-x} + 2^{2-x} + 2^{3-x} + 2^{4-x} + 2^{5-x} + 2^{6-x}]$
So, we consider the inequality

$$\begin{aligned} \frac{1}{6}(2^{-x})(2^1 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6) &\geq 1 \\ 21(2^{-x}) &\geq 1 \\ (2^{-x}) &\geq \frac{1}{21} \\ \log_2(2^{-x}) &\geq -\log_2(21) \\ -x &\geq -\log_2(21) \\ x &\leq \log_2(21) \end{aligned}$$

(b) For the following question assume that the ghosts have set the price of the game at $x = 4$. A fortune-teller is able to accurately predict whether the die roll will be even $(2, 4, 6)$ or odd $(1, 3, 5)$.

(i) [3 pts] Assume Pacman's utility is $U(r) = r$. The VPI (value of perfect information) of the prediction is:

● 0 ○ $\frac{1}{16}$ ○ $\frac{7}{8}$ ○ 1 ○ $\frac{7}{4}$

Because $x = 4$ and based on the answer for (ai), Pacman will not play the game and get a reward of 0. Hence, MEU($\emptyset$) = 0.
Now, let's calculate the MEU for the two predictions. Remember that Pacman has two choices, to play, or to stay (which gives him an utility of 0). Pacman will choose the best option.
MEU(fortune teller predicts odd) = $\max(0, \frac{1}{3}((1-4) + (3-4) + (5-4)) = 0$.
MEU(fortune teller predicts even) = $\max(0, \frac{1}{3}((2-4) + (4-4) + (6-4)) = 0$.
Hence, VPI(fortune teller prediction) = $\frac{1}{2}$ MEU(odd) + $\frac{1}{2}$ MEU(even) - MEU($\emptyset$) = 0

(ii) [3 pts] Assume Pacman's utility is $U'(r) = 2^r$. The VPI of the prediction is:

○ 0 ● $\frac{1}{16}$ ○ $\frac{7}{8}$ ○ 1 ○ $\frac{7}{4}$

From (aii), because $x = 4 \leq log_2(21)$, we know that Pacman will play the game, so MEU($\emptyset$) = EU(play). Hence, MEU($\emptyset$) = $\frac{1}{6}[2^{1-4} + 2^{2-4} + 2^{3-4} + 2^{4-4} + 2^{5-4} + 2^{6-4}] = \frac{1}{6}[\frac{1}{8} + \frac{1}{4} + \frac{1}{2} + 1 + 2 + 4] = \frac{21}{16}$.

Now, let's calculate the MEU for the two predictions. Remember that Pacman has two choices, to play, or to stay (which gives him an utility of 1). Pacman will choose the best option.
MEU(fortune teller predicts odd) = $\max(1, \frac{1}{3}(2^{1-4} + 2^{3-4} + 2^{5-4})) = \max(1, \frac{7}{8}) = 1$.
MEU(fortune teller predicts even) = $\max(1, \frac{1}{3}(2^{2-4} + 2^{4-4} + 2^{6-4})) = \max(1, \frac{7}{4}) = \frac{7}{4}$.
Hence, VPI(fortune teller prediction) = $\frac{1}{2}$ MEU(odd) + $\frac{1}{2}$ MEU(even) - MEU($\emptyset$) = $\frac{1}{2}(1) + \frac{1}{2}(\frac{7}{4}) - \frac{21}{16} = \frac{1}{16}$.

(c) [4 pts] For simplicity the following question concerns only Markov Decision Processes (MDPs) with no discounting ($\gamma = 1$) and parameters set up such that the total reward is always finite. Let J be the total reward obtained in the MDP:

$$J = \sum_{t=1}^{\infty} r(S_t, A_t, S_{t+1}).$$

The utility we've been using implicitly for MDPs with no discounting is $U(J) = J$. The value function $V(s)$ is equal to the maximum expected utility $E[U(J)] = E[J]$ if the start state is s, and it obeys the Bellman equation seen below:

$$V^*(s) = \max_a \sum_{s'} T(s, a, s')(r(s, a, s') + V^*(s')).$$

Now consider using the exponential utility $U'(J) = 2^J$ for MDPs. Write down the corresponding Bellman equation for $W^*(s)$, the maximum expected exponential utility $E[U'(J)] = E[2^J]$ if the start state is s.

$$W^*(s) = \max_a \sum_{s'} T(s, a, s') 2^{r(s,a,s')} W^*(s')$$

Let's parse the similarities between this solution and the regular Bellman update equation. We start at a maximizing node, s. From s, we have a set of actions we can take, and because we want to maximize our sum of utilities, we take a max over all the possible actions. This brings us to a chance node, $Q(s, a)$, which, branching from this node, contains many landing states. We land in one of the states s', based on the transition model $T(s, a, s')$.

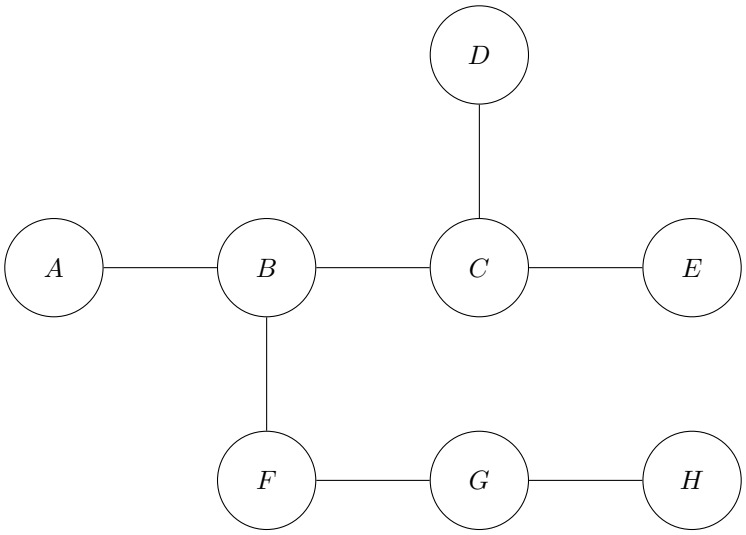
The difference lies in how we incorporate the utility of reward from the immediate rewards and the future rewards. Think back to the normal Bellman equation when $U(r) = r$. We receive an intermediate reward r, and add in the value at s', which corresponds to the sum of expected rewards acting optimally at s'. Let's say that the future reward sequence is R'. That means that when we append r to R', we get a new utility $U(r + R')$. But, knowing $U(r) = r$, then: $U(r + R') = U(r) + U(R')$, which is exactly the $r(s, a, s')$ and the $V^*(s')$ terms in the Bellman equation (with no discounting).

Now, when the utility function is $U(r) = 2^r$, then $U(r + R') = 2^{r+R'} = 2^r * 2^{R'}$. That is, the utilities of the immediate reward and the future rewards get multiplied together. This explains why the term $2^{r(s,a,s')}$ gets multiplied with $W^*(s')$ in the solutions.

Q7. [9 pts] Bayes Net CSPs

(a) For the following Bayes' Net structures that are missing a direction on their edges, assign a direction to each edge such that the Bayes' Net structure implies the requested conditional independences and such that the Bayes' Net structure does not imply the conditional independences requested not to be true. Keep in mind that Bayes' Nets cannot have directed cycles.

(i) [2 pts]

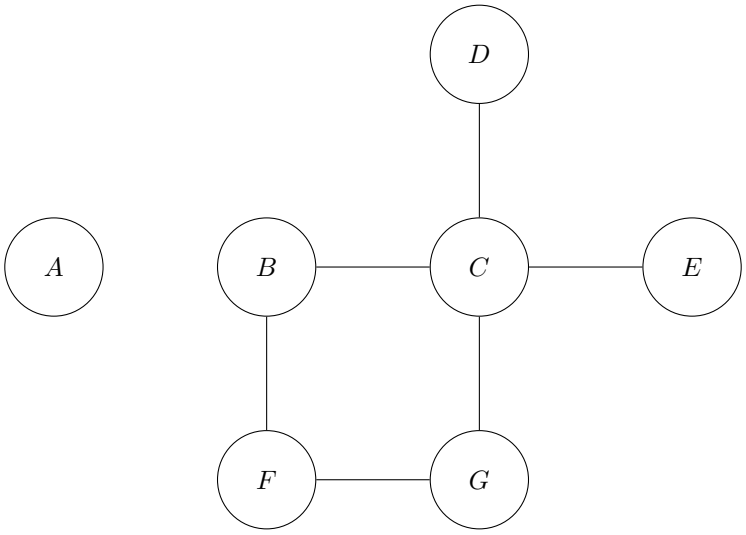


Constraints:

- $D \perp\!\!\!\perp G$
- not $D \perp\!\!\!\perp A$
- $D \perp\!\!\!\perp E$
- $H \perp\!\!\!\perp F$

The following are the directions of the edges:
$B \rightarrow A$
$C \rightarrow B$
$D \rightarrow C$
$E \rightarrow C$
$F \rightarrow B$
$F \rightarrow G$
$H \rightarrow G$

(ii) [2 pts]



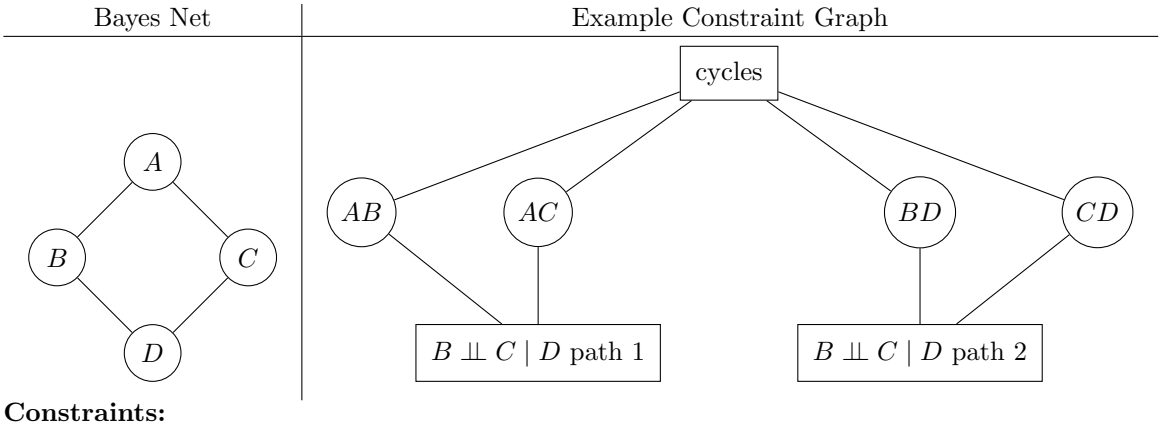
Constraints:

- $D \perp\!\!\!\perp F$
- not $D \perp\!\!\!\perp G$
- $D \perp\!\!\!\perp E$
- Bayes Net has no directed cycles

The following are the directions of the edges:
$C \to B$
$F \to B$
$F \to G$
$C \to G$
$D \to C$
$E \to C$

(b) For each of the following Bayes Nets and sets of constraints draw a constraint graph for the CSP. Remember that the constraint graph for a CSP with non-binary constraints, i.e., constraints that involve more than two variables, is drawn as a rectangle with the constraint connected to a node for each variable that participates in that constraint. A simple example is given below.

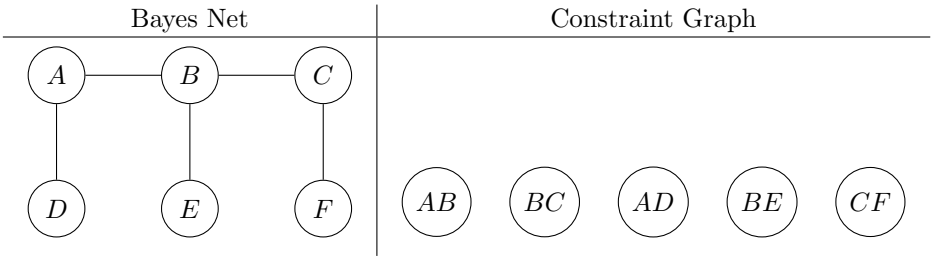
Note: As shown in the example below, if a constraint can be broken up into multiple constraints, do so.



Constraints:

- $B \perp\!\!\!\perp C \mid D$
- No directed cycles

(i) [2 pts]

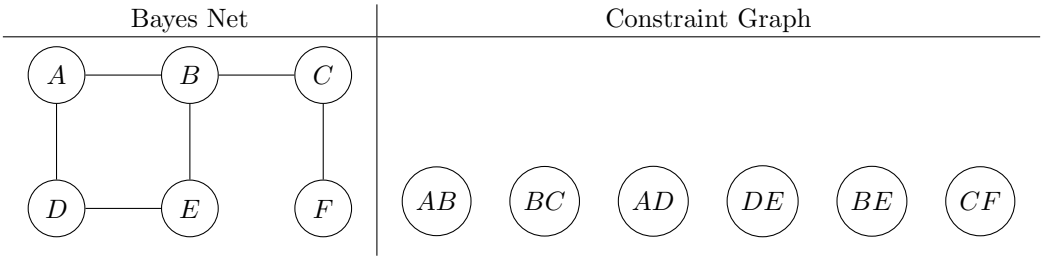


Constraints:

- $A \perp\!\!\!\perp F \mid E$
- not $D \perp\!\!\!\perp C$

Constraint $A \perp\!\!\!\perp F \mid E$: connect AB, BC, BE and CF.
Constraint not $D \perp\!\!\!\perp C$: connect AB, BC and AD.

(ii) [3 pts]



Constraints:

- $A \perp\!\!\!\perp E \mid F$

- $C \perp\!\!\!\perp E$
- No directed cycles

Constraint $A \perp\!\!\!\perp E \mid F$ with path going through path $A - B - E$ with descendant C and F: connect AB, BC, BE, CF.

Constraint $A \perp\!\!\!\perp E \mid F$ with path going through path $A - D - E$: connect AD, DE.

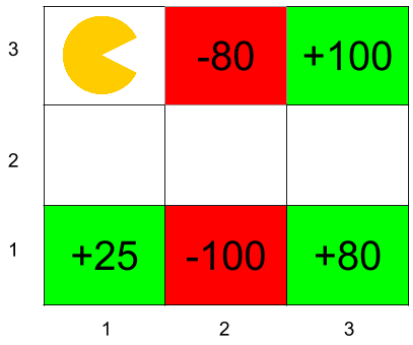
Constraint $C \perp\!\!\!\perp E$ with path going through path $C - B - E$: connect BC, BE.

Constraint $C \perp\!\!\!\perp E$ with path going through path $C - B - A - D - E$: connect AB, BC, AD, DE.

No direct cycles: connect AB, AD, DE and BE.

Q8. [8 pts] Q-Learning Strikes Back

Consider the grid-world given below and Pacman who is trying to learn the optimal policy. If an action results in landing into one of the shaded states the corresponding reward is awarded during that transition. All shaded states are terminal states, i.e., the MDP terminates once arrived in a shaded state. The other states have the *North, East, South, West* actions available, which deterministically move Pacman to the corresponding neighboring state (or have Pacman stay in place if the action tries to move out of the grad). Assume the discount factor $\gamma = 0.5$ and the Q-learning rate $\alpha = 0.5$ for all calculations. Pacman starts in state (1, 3).



(a) [2 pts] What is the value of the optimal value function V^* at the following states:

$V^*(3,2) =$ 100 $V^*(2,2) =$ 50 $V^*(1,3) =$ 12.5

The optimal values for the states can be found by computing the expected reward for the agent acting optimally from that state onwards. Note that you get a reward when you transition *into* the shaded states and not *out* of them. So for example the optimal path starting from (2,2) is to go to the +100 square which has a discounted reward of $0 + \gamma * 100 = 50$. For (1,3), going to either of +25 or +100 has the same discounted reward of 12.5.

(b) [3 pts] The agent starts from the top left corner and you are given the following episodes from runs of the agent through this grid-world. Each line in an Episode is a tuple containing (s, a, s', r).

Episode 1	Episode 2	Episode 3
(1,3), S, (1,2), 0	(1,3), S, (1,2), 0	(1,3), S, (1,2), 0
(1,2), E, (2,2), 0	(1,2), E, (2,2), 0	(1,2), E, (2,2), 0
(2,2), S, (2,1), -100	(2,2), E, (3,2), 0	(2,2), E, (3,2), 0
	(3,2), N, (3,3), +100	(3,2), S, (3,1), +80

Using Q-Learning updates, what are the following Q-values after the above three episodes:

$Q((3,2),\text{N}) =$ 50 $Q((1,2),\text{S}) =$ 0 $Q((2,2),E) =$ 12.5

Q-values obtained by Q-learning updates - $Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha(R(s,a,s') + \gamma \max_{a'} Q(s',a'))$.

(c) Consider a feature based representation of the Q-value function:

$$Q_f(s,a) = w_1 f_1(s) + w_2 f_2(s) + w_3 f_3(a)$$

$f_1(s)$: The x coordinate of the state $f_2(s)$: The y coordinate of the state

$$f_3(N) = 1,\ f_3(S) = 2,\ f_3(E) = 3,\ f_3(W) = 4$$

(i) [2 pts] Given that all w_i are initially 0, what are their values after the first episode:

$w_1 =$ -100 $w_2 =$ -100 $w_3 =$ -100

Using the approximate Q-learning weight updates: $w_i \leftarrow w_i + \alpha[(R(s,a,s') + \gamma \max_{a'} Q(s',a')) - Q(s,a)]f_i(s,a)$. The only time the reward is non zero in the first episode is when it transitions into the -100 state.

(ii) [1 pt] Assume the weight vector w is equal to $(1,1,1)$. What is the action prescribed by the Q-function in state $(2,2)$?

West

The action prescribed at (2,2) is $\max_a Q((2,2),a)$ where $Q(s,a)$ is computed using the feature representation. In this case, the Q-value for *West* is maximum $(2+2+4=8)$.

Q9. [11 pts] Probability and Bayes Nets

(a) [2 pts] Suppose $A \perp\!\!\!\perp B$. Determine the missing entries (x, y) of the joint distribution $P(A, B)$, where A and B take values in $\{0, 1\}$.

$$P(A = 0, B = 0) = 0.1$$
$$P(A = 0, B = 1) = 0.3$$
$$P(A = 1, B = 0) = x$$
$$P(A = 1, B = 1) = y$$

$x =$ ____.15____, $y =$ ____.45____

Note that $y/x = P(A = 1, B = 1)/P(A = 1, B = 0) = P(A = 0, B = 1)/P(A = 0, B = 0) = P(B = 1)/P(B = 0) = 3$ So $y = 3x$ and $x + y = 0.6$. Solve for x, y.

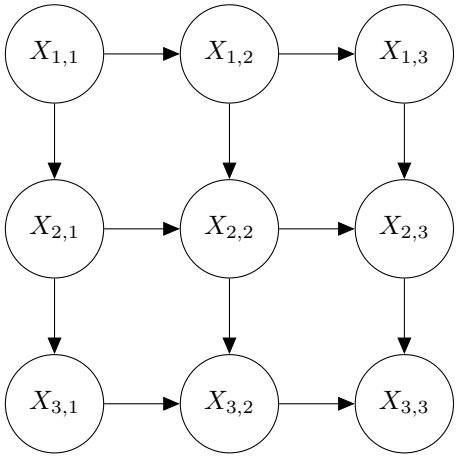
(b) [3 pts] Suppose $B \perp\!\!\!\perp C \mid A$. Determine the missing entries (x, y, z) of the joint distribution $P(A, B, C)$.

$$P(A = 0, B = 0, C = 0) = 0.01$$
$$P(A = 0, B = 0, C = 1) = 0.02$$
$$P(A = 0, B = 1, C = 0) = 0.03$$
$$P(A = 0, B = 1, C = 1) = x$$
$$P(A = 1, B = 0, C = 0) = 0.01$$
$$P(A = 1, B = 0, C = 1) = 0.1$$
$$P(A = 1, B = 1, C = 0) = y$$
$$P(A = 1, B = 1, C = 1) = z$$

$x =$ ____0.06____, $y =$ ____0.07____, $z =$ ____0.7____

First use the same observation about ratios as above to get that $x = 0.03 \cdot \frac{0.02}{0.01} = 0.06$. Then we have that $0.01 + 0.02 + 0.03 + 0.06 + 0.01 + 0.1 + y + z = 1$ so $y + z = 0.77$. The same observation about ratios gives $z/y = 10$. Solving, we get $y = 0.07, z = 0.7$.

(c) [3 pts] For this question consider the Bayes' Net below with 9 variables.



Which random variables are independent of $X_{3,1}$? (Leave blank if the answer is none.)

○ $X_{1,1}$ ○ $X_{1,2}$ ○ $X_{1,3}$ ○ $X_{2,1}$ ○ $X_{2,2}$ ○ $X_{2,3}$ ○ $X_{3,2}$ ○ $X_{3,3}$

There is at least one active path between $X_{3,1}$ and every other node.

Which random variables are independent of $X_{3,1}$ *given* $X_{1,1}$? (Leave blank if the answer is none.)

● $X_{1,2}$ ● $X_{1,3}$ ○ $X_{2,1}$ ○ $X_{2,2}$ ○ $X_{2,3}$ ○ $X_{3,2}$ ○ $X_{3,3}$

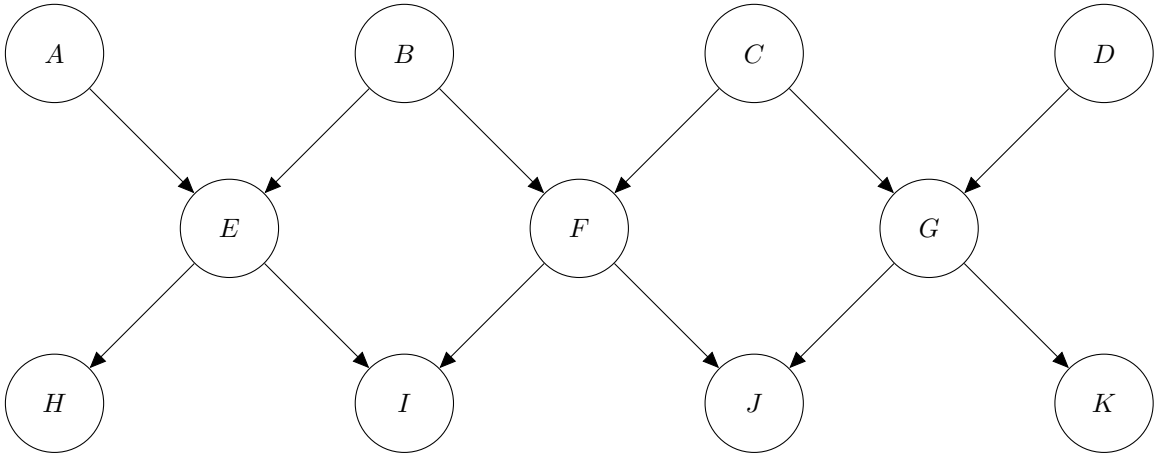
$X_{1,1}$ blocks the only active paths to both $X_{1,2}$ and $X_{1,3}$, so both of those become independent of $X_{3,1}$ given $X_{1,1}$

Which random variables are independent of $X_{3,1}$ *given* $X_{1,1}$ and $X_{3,3}$? (Leave blank if the answer is none.)

○ $X_{1,2}$ ○ $X_{1,3}$ ○ $X_{2,1}$ ○ $X_{2,2}$ ○ $X_{2,3}$ ○ $X_{3,2}$

The path from a node down to $X_{3,3}$ and up to another node is an active path.

For the following questions we will consider the following Bayes' Net:



(d) Consider a run of Gibbs sampling for the query $P(B, C \mid +h, +i, +j)$. The current sample value is $+a, +b, +c, +d, +e, +f, +g, +h, +i, +j, +k$. For each of the following scenarios, write out an expression for the distribution Gibbs sampling would sample from. *Your expression should contain only conditional probabilities available in the network, and your expression should contain a minimal number of such conditional probabilities.*

(i) [1 pt] If A were to be sampled next, the distribution over A to sample from would be:

$P(A| + b, +e) \propto P(+e|A, +b)P(+b)$

Note that only B, E are necessary because all the other variables are independent of A given B, E.

(ii) [1 pt] If F were to be sampled next, the distribution over F to sample from would be:

$P(F| + b, +c, +e, +g, +i, +j) \propto P(F| + b, +c)P(+i| + e, F)P(+j|F, +g)$

(iii) [1 pt] If K were to be sampled next, the distribution over K to sample from would be:

$P(K| + g)$