# CS 188 Fall 2015 — Introduction to Artificial Intelligence — HW8

**INSTRUCTIONS**

- **Due:** Wednesday, November 4, 2015 11:59 PM
- **Policy:** Can be solved in groups (acknowledge collaborators) but must be written up individually. However, we strongly encourage you to first work alone for about 30 minutes total in order to simulate an exam environment. Late homework will not be accepted.
- **Format:** Submit the answer sheet pdf containing your answers. You should solve the questions on this handout (either through a pdf annotator, or by printing, then scanning; we recommend the latter to match exam setting). **Make sure that your answers (typed or handwritten) are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.**
- **How to submit:** Go to www.gradescope.com. Log in and click on the class CS188 Fall 2015. Click on the submission titled HW 8 and upload your pdf containing your answers. If this is your first time using Gradescope, you will have to set your password before logging in the first time. To do so, click on "Forgot your password" on the login page, and enter your email address on file with the registrar's office (usually your @berkeley.edu email address). You will then receive an email with a link to reset your password.

| | |
|---|---|
| Last Name | |
| First Name | |
| SID | |
| Email | |
| Collaborators | |

**For staff use only**

| Q1 | Q2 | Q3 | Q4 | Total |
|---|---|---|---|---|
| /32 | /16 | /32 | /20 | /100 |

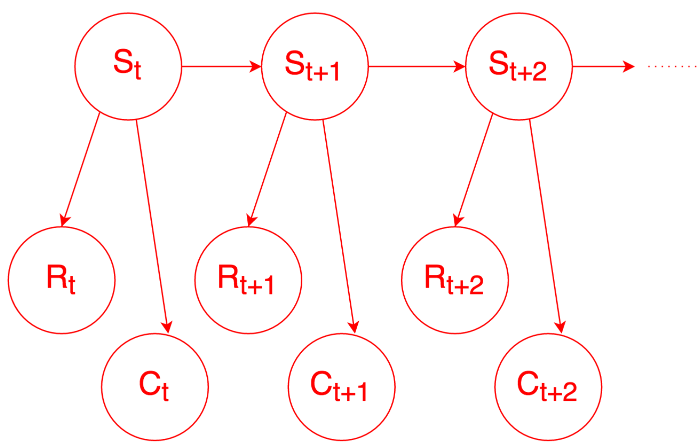
# Q1. [32 pts] Dynamic Bayes Net and Hidden Markov Model

A professor wants to know if students are getting enough sleep. Each day, the professor observes whether the students sleep in class, and whether they have red eyes. Let $S_t$ be the random variable of the student having enough sleep, $R_t$ be the random variable for the student having red eyes, and $C_t$ be the random variable of the student sleeping in class on day $t$. The professor has the following domain theory:

- The prior probability of getting enough sleep at time t, with no observations, is 0.6
- The probability of getting enough sleep on night $t$ is 0.9 given that the student got enough sleep the previous night, and 0.2 if not
- The probability of having red eyes is 0.1 if the student got enough sleep, and 0.7 if not
- The probability of sleeping in class is 0.2 if the student got enough sleep, and 0.4 if not

**(a)** [5 pts] The professor wants to formulate this information as a Dynamic Bayesian network. Provide a diagram and complete probability tables for the model.

**DBN diagram**




We will accept the answer that either takes $S_0$ or $S_1$ as prior.

**Probability Tables**

Again you can take either $S_0$ or $S_1$ as prior. If you assume $S_0$ for prior distribution, then we have the following probability tables

| $S_0$ | $P(S_0)$ |
|---|---|
| $s$ | 0.6 |
| $\neg s$ | 0.4 |

| $R_t$ | $S_t$ | $P(R_t \mid S_t)$ |
|---|---|---|
| $r$ | $s$ | 0.1 |
| $\neg r$ | $s$ | 0.9 |
| $r$ | $\neg s$ | 0.7 |
| $\neg r$ | $\neg s$ | 0.3 |

| $S_{t+1}$ | $S_t$ | $P(S_{t+1} \mid S_t)$ |
|---|---|---|
| $s_{t_1}$ | $s_t$ | 0.9 |
| $\neg s_{t_1}$ | $s_t$ | 0.1 |
| $s_{t_1}$ | $\neg s_t$ | 0.2 |
| $\neg s_{t_1}$ | $\neg s_t$ | 0.8 |

| $C_t$ | $S_t$ | $P(C_t \mid S_t)$ |
|---|---|---|
| $c$ | $s$ | 0.2 |
| $\neg c$ | $s$ | 0.8 |
| $c$ | $\neg s$ | 0.4 |
| $\neg c$ | $\neg s$ | 0.6 |

**(b)** [20 pts] Using DBN you defined and for the evidence values

$\neg r_1, \neg c_1$ = not red eyes, not sleeping in class

$r_2, \neg c_2$ = red eyes, not sleeping in class

$r_3, c_3$ = red eyes, sleeping in class

perform the following computations:

(i) State estimation: Compute $P(S_t|r_{1:t}, c_{1:t})$ for each of $t = 1, 2, 3$

$P(S_1|\neg r_1, \neg c_1)$ $S_0$ as prior : We will first compute $P(S_1)$ to get $P(S_1|\neg r_1, \neg c_1)$. This will correspond to predict and update step of our forward algorithm.

Predict : $P(s_1) = \sum_{s_0} P(s_1|s_0)P(s_0) = 0.6 * 0.9 + 0.4 * 0.2 = .62$

$$P(\neg s_1) = 1 - P(s_1) = .38$$

Update : $P(s_1|\neg r_1, \neg c_1) = \alpha P(\neg r_1, \neg c_1|s_1)P(s_1) = \alpha(0.9 * 0.8 * .62) = \alpha.4464$

$$P(\neg s_1|\neg r_1, \neg c_1) = \alpha P(\neg r_1, \neg c_1|\neg s_1)P(\neg s_1) = \alpha(0.3 * 0.6 * .38) = \alpha.0684$$

from these two we get, $\alpha = \frac{1}{.4464+.0684} = .5148$. So $P(s_1|\neg r_1, \neg c_1) = \frac{.4464}{.5148} = .867$, $P(\neg s_1|\neg r_1, \neg c_1) = \frac{.0684}{.5148} = .133$

If you assume $S_1$ to be your prior, then $P(S_1) = 0.6$ and $P(\neg S_1) = 0.4$. The observation at time 1 won't matter.

$P(S_2|r_{1:2}, c_{1:2})$ Again using forward algorithm.

Predict: $P(s_2|\neg r_1, \neg c_1) = \sum_{s_1} P(s_2|s_1)P(s_1|\neg r_1, \neg c_1) = .9 * .867 + .2 * .133 = .807$

$P(\neg s_2|\neg r_1, \neg c_1) = 1 - P(s_2|\neg r_1, \neg c_1) = .193$

Update: $P(s_2|r_{1:2}, c_{1:2}) = \alpha P(r_2, \neg c_2|s_2)P(s_2|\neg r_1, \neg c_1) = \alpha(.1 * .8 * .807) = \alpha * .06456$

$P(\neg s_2|r_{1:2}, c_{1:2}) = \alpha P(r_2, \neg c_2|\neg s_2)P(\neg s_2|\neg r_1, \neg c_1) = \alpha(.7 * .6 * .193) = \alpha * .08156$

$\rightarrow P(s_2|r_{1:2}, c_{1:2}) = .442, P(\neg s_2|r_{1:2}, c_{1:2}) = .558$

If you assume $S_1$ to be your prior, then $P(s_2|r_{1:2}, c_{1:2}) = .237, P(\neg s_2|r_{1:2}, c_{1:2}) = .763$

$P(S_3|r_{1:3}, c_{1:3})$ Predict: $P(s_3|r_{1:2}, c_{1:2}) = \sum_{s_2} P(s_3|s_2)P(s_2|r_{1:2}, c_{1:2}) = .9*.442+.2*.558 = .5094$

$P(\neg s_3|r_{1:2}, c_{1:2}) = 1 - P(s_3|r_{1:2}, c_{1:2}) = .4906$

Update: $P(s_3|r_{1:3}, c_{1:3}) = \alpha P(r_3, c_3|s_3)P(s_3|r_{1:2}, c_{1:2}) = \alpha(0.1 * .2 * .5094) = \alpha * .0102$

$P(\neg s_3|r_{1:3}, c_{1:3}) = \alpha P(r_3, c_3|\neg s_3)P(\neg s_3|r_{1:2}, c_{1:2}) = \alpha(.7 * .4 * .4906) = \alpha * .1374$

$\rightarrow P(s_3|r_{1:3}, c_{1:3}) = .069, P(\neg s_3|r_{1:3}, c_{1:3}) = .931$

If you assume $S_1$ to be your prior, then $P(s_3|r_{1:3}, c_{1:3}) = .0396, P(\neg s_3|r_{1:3}, c_{1:3}) = .9604$

(ii) Smoothing: Compute $P(S_t|r_{1:3}, c_{1:3})$ for $t = 2, 3$ (Hint: AIMA pg. 574)

$P(S_3|r_{1:3}, c_{1:3})$ same as part (a)

$P(s_3|r_{1:3}, c_{1:3}) = .069, P(\neg s_3|r_{1:3}, c_{1:3}) = .931$

or $P(s_3|r_{1:3}, c_{1:3}) = .0396, P(\neg s_3|r_{1:3}, c_{1:3}) = .9604$

$P(S_2|r_{1:3}, c_{1:3})$ First compute backward message $P(r_3, c_3|S_2)$.

$$P(r_3, c_3|S_2) = \sum_{s_3} P(r_3, c_3|s_3)P(s_3|S_2) \rightarrow P(r_3, c_3|s_2) = .02*.9+.28*.1 = .046, P(r_3, c_3|\neg s_2) = .228$$

$$P(S_2|r_{1:3}, c_{1:3}) = \alpha P(S_2|r_{1:2}, c_{1:2})P(r_3, c_3|S_2)$$

$$\rightarrow P(s_2|r_{1:3}, c_{1:3}) = \alpha.442 * .046 = \alpha * .0203, P(\neg s_2|r_{1:3}, c_{1:3}) = \alpha.558 * .228 = .127$$

$$\rightarrow P(s_2|r_{1:3}, c_{1:3}) = .138, P(\neg s_2|r_{1:3}, c_{1:3}) = .862$$

or $P(s_2|r_{1:3}, c_{1:3}) = .059, P(\neg s_2|r_{1:3}, c_{1:3}) = .941$

At every time step t for t = 1 to n, you observe a tuple $(r_t, c_t)$ telling you whether the student had red eyes and whether they were sleeping in class. Given these observations and $P(S_k|r_{1:k}, c_{1:k})$, find an expression for $P(S_k|r_{1:n}, c_{1:n})$, where $0 \leq k \leq n$. You may only use the probability tables in the DBN and $P(S_k|r_{1:k}, c_{1:k})$

$P(S_k|r_{1:n}, c_{1:n})$

$$\begin{aligned} P(S_k|r_{1:n}, c_{1:n}) &= P(S_k|r_{1:k}, c_{1:k}, r_{k+1:n}, c_{k+1:n}) \\ &= \alpha P(S_k|r_{1:k}, c_{1:k})P(r_{k+1:n}, c_{k+1:n}|S_k, r_{1:k}, c_{1:k}) \\ &= \alpha P(S_k|r_{1:k}, c_{1:k})P(r_{k+1:n}, c_{k+1:n}|S_k) \end{aligned}$$

We can derive backward message $P(r_{k+1:n}, c_{k+1:n}|S_k)$ by the following recursive formula

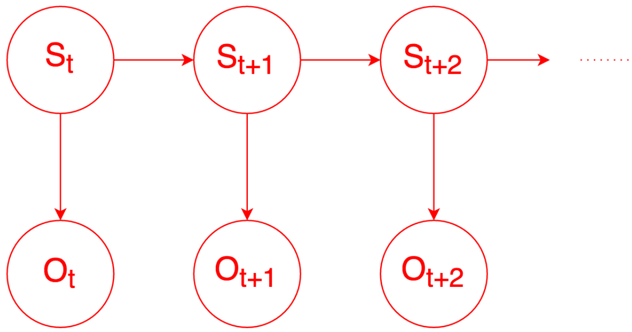
$$\begin{aligned} P(r_{k+1:n}, c_{k+1:n}|S_k) &= \sum_{s_{k+1}} P(r_{k+1:n}, c_{k+1:n}, s_{k+1}|S_k) \\ &= \sum_{s_{k+1}} P(r_{k+1:n}, c_{k+1:n}|s_{k+1}, S_k)P(s_{k+1}|S_k) \\ &= \sum_{s_{k+1}} P(r_{k+1}, c_{k+1}, r_{k+2:n}, c_{k+2:n}|s_{k+1})P(s_{k+1}|S_k) \\ &= \sum_{s_{k+1}} P(r_{k+1}|s_{k+1})P(c_{k+1}|s_{k+1})P(r_{k+2:n}|s_{k+1})P(c_{k+2:n}|s_{k+1})P(s_{k+1}|S_k) \end{aligned}$$

So the final form for general smoothing formula is,

$$P(S_k|r_{1:n}, c_{1:n}) = \alpha P(S_k|r_{1:k}, c_{1:k}) \sum_{s_{k+1}} P(r_{k+1}|s_{k+1})P(c_{k+1}|s_{k+1})P(r_{k+2:n}|s_{k+1})P(c_{k+2:n}|s_{k+1})P(s_{k+1}|S_k)$$

**(c)** [5 pts] Reformulate the problem as a hidden Markov model. Provide a diagram and complete probability tables for the model.

**HMM diagram**




Again we will accept a diagram with $S_0$ or $S_1$

**Probability Tables**

Instead of two observation variables in DBN, we will shrink two observations to one. This new observation variables $O_t$ can take four values. $S_0$ can be $S_1$

| $S_0$ | $P(S_0)$ |
|---|---|
| $s$ | 0.6 |
| $\neg s$ | 0.4 |

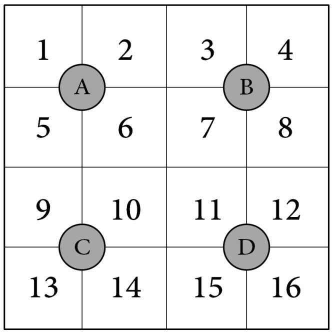
| $S_{t+1}$ | $S_t$ | $P(S_{t+1}\|S_t)$ |
|---|---|---|
| $s_{t_1}$ | $s_t$ | 0.9 |
| $\neg s_{t_1}$ | $s_t$ | 0.1 |
| $s_{t_1}$ | $\neg s_t$ | 0.2 |
| $\neg s_{t_1}$ | $\neg s_t$ | 0.8 |

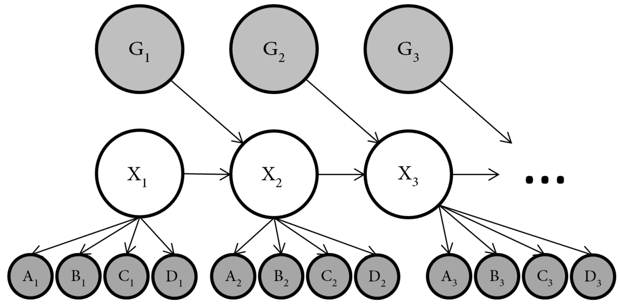
| $O_t$ | $S_t$ | $P(O_t\|S_t)$ |
|---|---|---|
| $r, c$ | $s$ | 0.02 |
| $r, \neg c$ | $s$ | 0.08 |
| $\neg r, c$ | $s$ | 0.18 |
| $\neg r, \neg c$ | $s$ | 0.72 |
| $r, c$ | $\neg s$ | 0.28 |
| $r, \neg c$ | $\neg s$ | 0.42 |
| $\neg r, c$ | $\neg s$ | 0.12 |
| $\neg r, \neg c$ | $\neg s$ | 0.18 |

**(d)** [2 pts] **True/False:** The probabilities computed in part (b) would have been different if you used the HMM instead of the DBN.

# Q2. [16 pts] Finding WALL-E

We would like to track the location of our friendly garbage collecting robot, WALL-E. WALL-E lives in a 4x4 Manhattan grid city, as shown below. The structure of the DBN is given below, which includes $X$, the position of WALL-E; $G$, the readings from a garbage sensor; and $(A,B,C,D)$, readings from the motion sensors.




The garbage sensor $G$ takes on a value in $\{1, 2, ..., 16\}$ corresponding to the square with the most garbage at time $t$. WALL-E is programmed to move toward the square with the most garbage, but he will only take an optimal action with probability 0.8. In each time step, WALL-E can either stay in the same square, or he can move to an adjacent square. In the case where multiple actions would move you equally close to the desired position, WALL-E has an equal probability of taking any of these actions. In the case that WALL-E fails to take an optimal action, he has an equal probability of taking any of the non-optimal actions.

For example, if $X_t = 2$ and $G_t = 15$, the transition model will look like this:

| $X_{t+1}$ | $P(X_{t+1} \mid X_t = 2, G_t = 15)$ |
|---|---|
| 1 | 0.1 |
| 2 | 0.1 |
| 3 | 0.4 |
| 6 | 0.4 |

The motion sensors, ($A$, $B$, $C$, $D$), take on a value in $\{ON, OFF\}$. At a time $t$, the sensor adjacent to the square that WALL-E is on always outputs $ON$. Otherwise, the sensor will output $ON$ or $OFF$ with probability 0.4 and 0.6. For example, the sensor tables would look like this if $X = 6$

| $A$ | $P(A \mid X = 6)$ |
|---|---|
| $ON$ | 1 |
| $OFF$ | 0 |

| $B$ | $P(B \mid X = 6)$ |
|---|---|
| $ON$ | 0.4 |
| $OFF$ | 0.6 |

| $C$ | $P(C \mid X = 6)$ |
|---|---|
| $ON$ | 0.4 |
| $OFF$ | 0.6 |

| $D$ | $P(D \mid X = 6)$ |
|---|---|
| $ON$ | 0.4 |
| $OFF$ | 0.6 |

**(a)** [6 pts]

Let's say the initial particles you have are $[X_t = 2, X_t = 12, X_t = 13]$. You get the following readings from your sensors $[A = ON, B = OFF, C = ON, D = OFF, G_{t-1} = 2]$.

What is the weight for each particle?

| Particle | Weight |
|---|---|
| $X_t = 2$ | P(A=ON\|X=2)P(B=OFF\|X=2)P(C=ON\|X=2)P(D=OFF\|X=2) = (1)(0.6)(0.4)(0.6) = 0.144 |
| $X_t = 12$ | P(A=ON\|X=12)P(B=OFF\|X=12)P(C=ON\|X=12)P(D=OFF\|X=12) = (0.4)(0.6)(0.4)(0) = 0 |
| $X_t = 13$ | P(A=ON\|X=13)P(B=OFF\|X=13)P(C=ON\|X=13)P(D=OFF\|X=13) = (0.4)(0.6)(1)(0.6) = 0.144 |

The value of $G_{t-1}$ does not affect the value of each weight

**(b)** [6 pts] It seems, much to your dismay, that sensor $C$ is broken, and will always give a reading of $ON$. Recalculate the weights with this new knowledge.

| Particle | Weight |
|---|---|
| $X_t = 2$ | P(A=ON\|X=2)P(B=OFF\|X=2)P(C=ON\|X=2)P(D=OFF\|X=2) = (1)(0.6)(1)(0.6) = 0.36 |
| $X_t = 12$ | P(A=ON\|X=12)P(B=OFF\|X=12)P(C=ON\|X=12)P(D=OFF\|X=12) = (0.6)(0.4)(1)(0) = 0 |
| $X_t = 13$ | P(A=ON\|X=13)P(B=OFF\|X=13)P(C=ON\|X=13)P(D=OFF\|X=13) = (0.4)(0.6)(1)(0.6) = 0.144 |

**(c)** [4 pts] To decouple this question from the previous question, let's say that the weights you found for each particle are as follows.

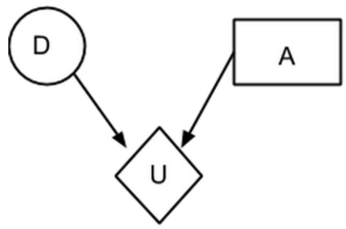
| Particle | Weight |
|---|---|
| $X_t = 8$ | 0.24 |
| $X_t = 14$ | 0.1 |
| $X_t = 11$ | 0.16 |

If you were to resample 100 new particles, what is the expected number of particles that will be $X = 11$?

Expected number of particles = $\frac{0.16}{0.24+0.1+0.16} \times 100 = 32$

# Q3. [32 pts] Bayes' Nets and Decision Networks

It is Monday night, and Bob is finishing up preparing for the CS188 Midterm. Bob has already mastered all the topics except one: Decision Networks. He is contemplating whether to spend the remainder of his evening reviewing that topic (*review*), or just go to sleep (*sleep*). Decision Networks are either going to be on the test ($d$) or not be on the test ($\neg d$). His utility of satisfaction is only affected by these two variables as shown below:




| D | P(D) |
|---|---|
| $d$ | 0.6 |
| $\neg d$ | 0.4 |

| D | A | U(D,A) |
|---|---|---|
| $d$ | *review* | 1200 |
| $\neg d$ | *review* | 400 |
| $d$ | *sleep* | 0 |
| $\neg d$ | *sleep* | 1600 |

**(a)** [6 pts] **Maximum Expected Utility**

Compute the following quantities:

EU($review$) =

$$P(d)U(d|review) + P(\neg d)U(\neg d, review) = 0.6 * 1200 + 0.4 * 400 = 880$$

$EU(sleep)$ =

$$P(d)U(d, sleep) + P(\neg d)U(\neg d, sleep) = 0.6 * 0 + 0.4 * 1600 = 640$$

$MEU(\{\})$ =

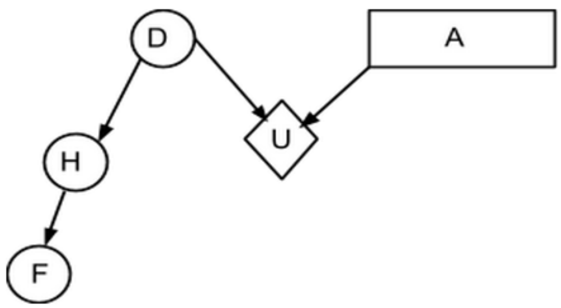
$$max(880, 640) = 880$$

Action that achieves $MEU(\{\})$ = $review$

This result notwithstanding, you should get some sleep.

**(b)** [14 pts] **The TA is on Facebook**

The TAs happiness ($H$) is affected by whether decision networks are going to be on the exam. The happiness ($H$) determines whether the TA posts on Facebook ($f$) or doesn't post on Facebook ($\neg f$). The prior on $D$ and utility tables remain unchanged.




| F | H | $P(F\|H)$ |
|---|---|---|
| $f$ | $h$ | 0.6 |
| $\neg f$ | $h$ | 0.4 |
| $f$ | $\neg h$ | 0.2 |
| $\neg f$ | $\neg h$ | 0.8 |

| H | D | $P(H\|D)$ |
|---|---|---|
| $h$ | $d$ | 0.95 |
| $\neg h$ | $d$ | 0.05 |
| $h$ | $\neg d$ | 0.25 |
| $\neg h$ | $\neg d$ | 0.75 |

| D | P(D) |
|---|---|
| $d$ | 0.6 |
| $\neg d$ | 0.4 |

| D | A | U(D,A) |
|---|---|---|
| $d$ | *review* | 1200 |
| $\neg d$ | *review* | 400 |
| $d$ | *sleep* | 0 |
| $\neg d$ | *sleep* | 1600 |

Decision network.

Tables that define the model are shown above.

| H | $P(H)$ |
|---|---|
| $h$ | 0.67 |
| $\neg h$ | 0.33 |

| F | $P(F)$ |
|---|---|
| $f$ | 0.468 |
| $\neg f$ | 0.532 |

| D | F | $P(D\|F)$ |
|---|---|---|
| $d$ | $f$ | 0.744 |
| $\neg d$ | $f$ | 0.256 |
| $d$ | $\neg f$ | 0.474 |
| $\neg d$ | $\neg f$ | 0.526 |

| F | D | $P(F\|D)$ |
|---|---|---|
| $f$ | $d$ | 0.58 |
| $\neg f$ | $d$ | 0.42 |
| $f$ | $\neg d$ | 0.3 |
| $\neg f$ | $\neg d$ | 0.7 |

| D | H | $P(D\|H)$ |
|---|---|---|
| $d$ | $h$ | 0.85 |
| $\neg d$ | $h$ | 0.15 |
| $d$ | $\neg h$ | 0.09 |
| $\neg d$ | $\neg h$ | 0.91 |

Tables computed from the first set of tables. Some of them might be convenient to answer the questions below.

Compute the following quantities:

$EU(review|f) = P(d|f)U(d, review) + P(\neg d|f)U(\neg d, review) = 0.744 * 1200 + 0.256 * 400 = 995.2$

$EU(sleep|f) = P(d|f)U(d, sleep) + P(\neg d|f)U(\neg d, sleep) = 0.744 * 0 + 0.256 * 1600 = 409.6$

$MEU(\{f\}) = \max(995.2, 409.6) = 995.2$

Optimal Action($\{f\}$) = *review*

$EU(review|\neg f) = P(d|\neg f)U(d, review) + P(\neg d|\neg f)U(\neg d, review) = 0.474 * 1200 + 0.526 * 400 = 779.2$

$EU(sleep|\neg f) = P(d|\neg f)U(d, sleep) + P(\neg d|\neg f)U(\neg d, sleep) = 0.474 * 0 + 0.526 * 1600 = 841.6$

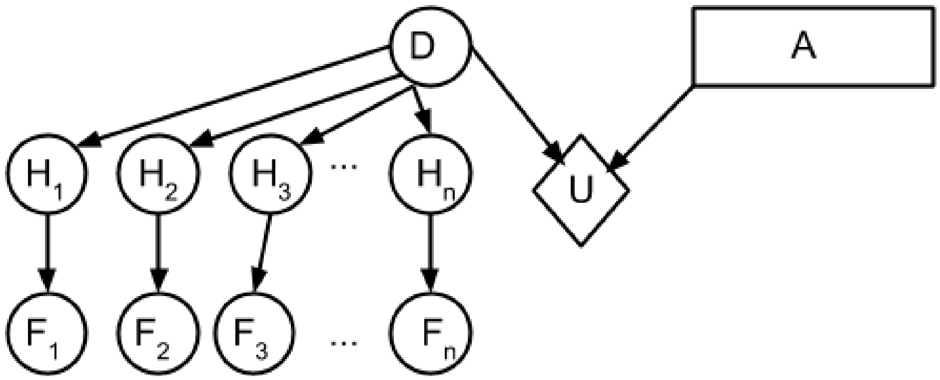
$MEU(\{\neg f\}) = \max(779.2, 841.6) = 841.6$

Optimal Action($\{\neg f\}$) = *sleep*

$VPI(\{F\}) = P(f)MEU(\{f\}) + P(\neg f)MEU(\{\neg f\}) - MEU(\{\}) = 0.468 * 995.2 + .532 * 841.6 - 880 = 33.48$

**(c)** [12 pts] **VPI Comparisons**

Now consider the case where there are $n$ TAs. Each TA follows the same probabilistic models for happiness ($H$) and posting on Facebook ($F$) as in the previous question.




True False $VPI(H_1|F_1) = 0$

Justify: $F_1$ is just a noisy version of $H_1$. Hence finding out $H_1$ gives us more information about $D$ even when we have already observed $F_1$. This in turn will allow us to more often make the right decision between *sleep* and *review*.

True False $VPI(F_1|H_1) = 0$

Justify: The parent variable of the utility node, $D$, is conditionally independent of $F_1$ given $H_1$.

True False $VPI(F_3|F_2, F_1) > VPI(F_2|F_1)$

Justify: The $F_i$ variables give us noisy information about $D$. The more $F_i$ variables we get to observe, the better chance we end up being able to make the right decision. The more $F_i$ variables we have already observed, however, the less an additiona observation of a new variable $F_j$ will influence the distribution of $D$.

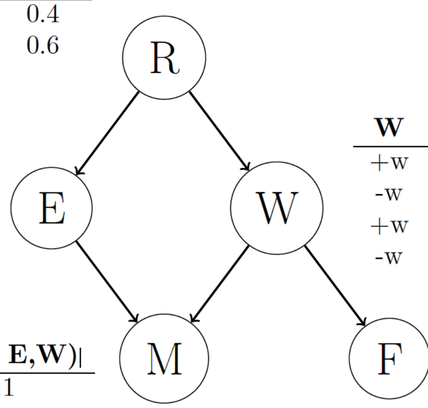
True False $VPI(F_1, F_2, \ldots, F_n) < VPI(H_1, H_2, \ldots, H_n)$

Justify: The $F_i$ variables are noisy versions of the $H_i$ variables, hence observing the $H_i$ variables is more valuable.

# Q4. [20 pts] Sampling

Consider the following Bayes Net and corresponding probability tables.

| R | P(R) |
|---|---|
| +r | 0.4 |
| -r | 0.6 |




| E | R | P(E \| R) |
|---|---|---|
| +e | +r | 0.3 |
| -e | +r | 0.7 |
| +e | -r | 0.6 |
| -e | -r | 0.4 |

| W | R | P(W \| R) |
|---|---|---|
| +w | +r | 0.9 |
| -w | +r | 0.1 |
| +w | -r | 0.2 |
| -w | -r | 0.8 |

| M | E | W | P(M \| E,W) |
|---|---|---|---|
| +m | +e | +w | 0.1 |
| -m | +e | +w | 0.9 |
| +m | +e | -w | 0.45 |
| -m | +e | -w | 0.55 |
| +m | -e | +w | 0.35 |
| -m | -e | +w | 0.65 |
| +m | -e | -w | 0.9 |
| -m | -e | -w | 0.1 |

| F | W | P(F \| W) |
|---|---|---|
| +f | +w | 0.15 |
| -f | +w | 0.85 |
| +f | -w | 0.75 |
| -f | -w | 0.25 |

We are going to use sampling to approximate the query $P(R|f,m)$. We have the following 3 samples.

$$(r, e, \neg w, m, f) \quad (r, \neg e, w, \neg m, f) \quad (r, e, \neg w, m, f)$$

**(a)** [12 pts] **Probability:** Fill in the following table with the probabilities of drawing each respective sample given that we are using each of the following sampling techniques. (Hint: $P(f,m) = .181$)

| $P(sample \mid method)$ | $(r, e, \neg w, m, f)$ | $(r, \neg e, w, \neg m, f)$ |
|---|---|---|
| prior sampling | $.4 * .3 * .1 * .45 * .75 = .00405$ | $.4 * .7 * .9 * .65 * .15 = 0.02457$ |
| rejection sampling | $\frac{P(r,e,\neg w,m,f)}{P(m,f)} = \frac{.00405}{.181} = .0224$ | 0 |
| likelihood weighting | $P(r)P(e\|r)P(\neg w\|r) = .4 * .3 * .1 = .012$ | 0 |

**(b)** [8 pts] We are going to use Gibbs sampling to estimate the probability of getting the third sample $(r, e, \neg w, m, f)$. We will start from the sample $(\neg r, \neg e, \neg w, m, f)$ and resample E first then R. What is the probability of drawing sample $(r, e, \neg w, m, f)$?

**Put your answer here:**

$$P(e|\neg r, \neg w, m, f) = \frac{P(\neg w, \neg r, e, m, f)}{\sum_e P(\neg w, \neg r, e, m, f)} = \frac{.6 * .6 * .8 * .45 * .75}{.6 * .6 * .8 * .45 * .75 + .6 * .4 * .8 * .9 * .75} = .4286$$

$$P(r|e, \neg w, m, f) = \frac{P(r, e, \neg w, m, f)}{\sum_r P(r, e, \neg w, m, f)} = \frac{.4 * .3 * .1 * .45 * .75}{.4 * .3 * .1 * .45 * .75 + .6 * .6 * .8 * .45 * .75} = .04$$

The probability of sampling $(\neg r, \neg e, w, m, f)$ is the product of three sampling probabilities. So .4286 * .04 = 0.017.