Q1. Particle Filtering

This pseudocode outlines the particle filtering algorithm discussed in lecture. The variable x represents a list of N particles, while w is a list of N weights for those particles.

```
1: procedure PARTICLE FILTERING(T, N)       ▷ T: number of time steps, N: number of sampled particles
2:     x ← sample N particles from initial state distribution P(X_0)                  ▷ Initialize
3:     for t ← 0 to T − 1 do                 ▷ X_t: hidden state, E_t: observed evidence
4:         x_i ← sample particle from P(X_{t+1}|X_t = x_i) for i = 1, ..., N          ▷ Time Elapse Update
5:         w_i ← P(E_{t+1}|X_{t+1} = x_i) for i = 1, ..., N                           ▷ Evidence Update
6:         x ← resample N particles according to weights w                            ▷ Particle Resampling
7:     return x
```

1. Here, we consider the unweighted particles in x as approximating a distribution.

   (a) After executing line 2, which distribution do the particles x represent?

   ○ $P(X_{t+1}|E_{1:t})$ ○ $P(X_0)$ ○ $P(X_{t+1}|E_{1:t+1})$ ○ None

   (b) After executing line 5, which distribution do the weights w **alone** represent?

   ○ $P(X_{t+1}|X_t, E_{1:t+1})$ ○ $P(X_{1:t+1}|E_{1:t+1})$ ○ $P(X_{t+1}|E_{1:t+1})$ ○ None

2. The particle filtering algorithm should return a sample-based approximation to the true posterior distribution $P(X_T \mid E_{1:T})$. The algorithm is **consistent** if and only if the approximation converges to the true distribution as $N \to \infty$. In this question, we present several modifications to Algorithm **??**. For each modification, indicate if the algorithm is still **consistent** or **not consistent**, and if it is **consistent**, indicate whether you expect it to be **more accurate** in general in terms of its estimate of $P(X_T \mid E_{1:T})$ (i.e., you would expect the estimated distribution to be closer to the true one) or **less accurate**. *Assume unlimited computational resources and arbitrary precision arithmetic.*

   (a) Replace lines 4–6 as follows:
   $4'$: Compute $P(X_t = q|E_{1:t})$, for all states q.
   $5'$: Exactly calculate $P(X_{t+1}|E_{1:t+1})$.
   $6'$: Resample N particles according to $5'$. Assign to x.
   This algorithm is:

   ○ Consistent and More Accurate ○ Not consistent
   ○ Consistent and Less Accurate

   Briefly explain why:

   (b) We modify line 6 to uniformly sample a random number of particles in the range $[1, N]$ according to the weights at each time step (as opposed to a fixed number of particles N at every time step). You can assume that the sensor model is not deterministic. You should consider the limit as both $T \to \infty$ and $N \to \infty$ in your analysis rather than letting T be fixed. This algorithm is:

   ○ Consistent and More Accurate ○ Not Consistent
   ○ Consistent and Less Accurate

   Briefly explain why:

(c) We initialize each w_i to 1 before entering the for loop. Replace lines 5 and 6 with the following update. Notice that line 6 is now outside of the loop.

5′ $w_i \leftarrow w_i + \log_2(P(E_{t+1}|X_{t+1} = x_i))$ for $i = 1, \ldots, N$.

6′ $x \leftarrow$ resample N particles according to the exponential of the weights, 2^w.

This algorithm is:

- ○ Consistent and More Accurate
- ○ Consistent and Less Accurate
- ○ Not Consistent

Briefly explain why:

Q2. Probability

(a) $\{V, W, X, Y, Z\}$ are random variables. V has domain $\{v_1, v_2, v_3\}$ while the rest are boolean random variables.

(i) Suppose we were to write out the full probability tables for the following distributions. Specify the number of entries (cells) that would be in the corresponding probability table and the sum of the probabilities in the table. Write "-" if the answer cannot be determined.

Table	Entries	Sum
$\mathbf{P}(X \mid V = v, Z)$		
$\mathbf{P}(V = v, W, X \mid Y, Z)$		
$\mathbf{P}(X, Y \mid V)$		
$\mathbf{P}(X, Y, Z \mid W = w)$		

(ii) **What is the fewest** number of parameters (individual probabilities) required to fully specify the distribution $\mathbf{P}(W \mid V, Y = y, Z)$?
(By specify, we mean provide enough information to write out the full table)

(b) Consider the different ways of factoring the joint distribution $\mathbf{P}(J, K, L)$ using the chain rule, **without making any independence assumptions**. Check the boxes for **any** expression that is a valid factoring:

☐ $\mathbf{P}(J \mid K)\, \mathbf{P}(K \mid L)\, \mathbf{P}(L)$

☐ $\mathbf{P}(K, L \mid J)\, \mathbf{P}(J)$

☐ $\mathbf{P}(J \mid K, L)\, \mathbf{P}(K)\, \mathbf{P}(L)$

☐ $\mathbf{P}(L \mid J)\, \mathbf{P}(J \mid K)\, \mathbf{P}(K)$

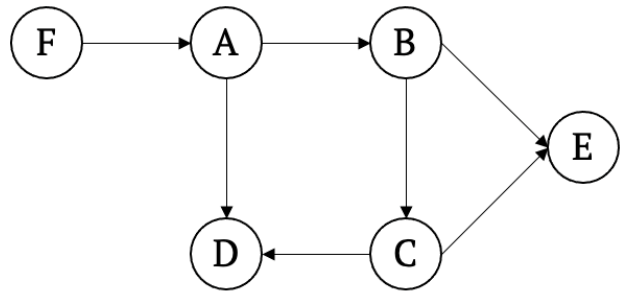
☐ $\mathbf{P}(L \mid K, J)\, \mathbf{P}(K \mid J)\, \mathbf{P}(J)$

○ None

Q3. Independence and D-Separation

(a) Conditional Independence

Consider the following Bayes' net:



In our notation, $G \perp\!\!\!\perp H$ means "G is independent of H given nothing", and $G \perp\!\!\!\perp H \mid \{I, J\}$ means "G is independent of H given I and J."

Indicate whether each of the claims below are True or False by filling in the corresponding circle. If you indicate False, also give an example active triple along the path, e.g. $Y \rightarrow U \rightarrow M$. It's the order that matters so Y, U, M and M, U, Y are also fine.

(i) It is guaranteed that $D \perp\!\!\!\perp B$

○ True ○ False Triple: ☐

(ii) It is guaranteed that $A \perp\!\!\!\perp C \mid B$

○ True ○ False Triple: ☐

(iii) It is guaranteed that $A \perp\!\!\!\perp E \mid C$

○ True ○ False Triple: ☐

(iv) It is guaranteed that $F \perp\!\!\!\perp B \mid \{A, C\}$.

○ True ○ False Triple: ☐

(v) It is guaranteed that $F \perp\!\!\!\perp C \mid E$

○ True ○ False Triple: ☐

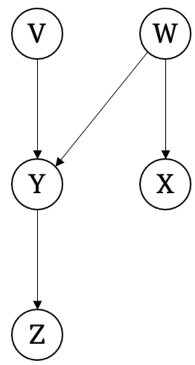
(vi) It is guaranteed that $F \perp\!\!\!\perp E \mid \{B, D\}$

○ True ○ False Triple: ☐

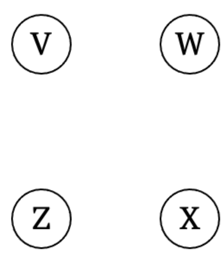
(b) Marginalization and Conditioning

Consider a Bayes' net encompassing the random variables V, W, X, Y, Z with the structure shown. Note the full joint distribution $\mathbf{P}(V, W, X, Y, Z)$. In each of these questions think about which conditional independence relations the Bayes' Nets must encode.

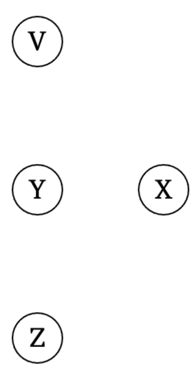
The following questions are unrelated, meaning your answer to part(i) should not influence part(ii).



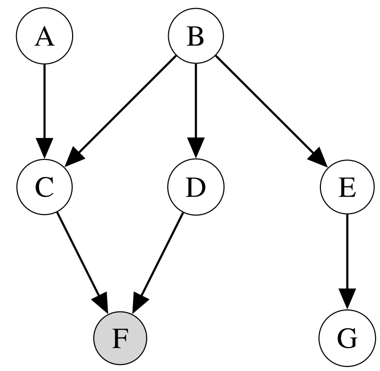
(i) Consider the marginal distribution $\mathbf{P}(V, W, X, Z) = \sum_y \mathbf{P}(V, W, X, Y = y, Z)$, where Y was eliminated. On the diagram below, draw the Bayes' net structure which represents this marginal distribution using the **fewest** arrows possible. Write *"No arrows needed"* if arrows are not needed.



(ii) Assume we are given this observation: $W = w$. On the diagram below, draw Bayes' net structure which represents the conditional distribution $\mathbf{P}(V, X, Y, Z)$ using the **fewest** arrows possible. If no arrows are needed write *"No arrows needed."*



Q4. Variable Elimination



Bayes' Net **N**

In this problem, carry out variable elimination inference for $\mathbf{P}(G|+f)$ in the Bayes' net N (pictured on the right). All variables have binary domains. Eliminate variables in the order prescribed below.

Answer format: when writing factors made during elimination, include the factor number, arguments, elimination sum, and joined factors like so:

$$f_1(X,+y) = \sum_z P(z|+y)P(X|z)$$

(a) What factors do we start with after incorporating evidence?

(b) Eliminate A to make a new factor f_1:

(c) Eliminate B to make a new factor f_2:

(d) Eliminate C to make a new factor f_3:

(e) Eliminate E to make a new factor f_4:

(f) Eliminate D to make a new factor f_5:

(g) Compute $P(+g|+f)$ from f_5. (Normalize f_5). Your answer should only be in terms of instantiations of f_5.

$P(+g|+f) =$

(h) **(i)** Amongst $f_1, f_2, \ldots, f_5$, which is the largest factor generated? (Assume all variables have binary domains)

(ii) How large is this factor? (Assume all variables have binary domains)

Q5. Hours Worked

(a) How many hours did you spend on this homework? Any reasonable answer (number greater than zero) will receive credit. This will not affect your score on any other problem.