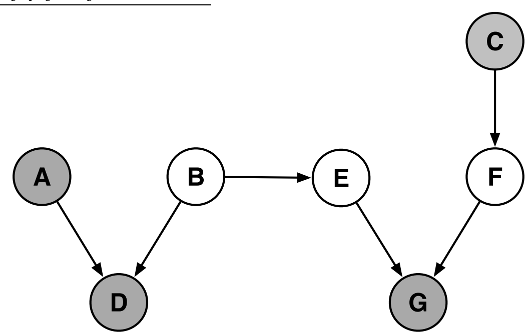
**CSE 150. Assignment 3** *Spring 2018*

**Out:** *Tue Apr 24*
**Due:** *Tue May 01*

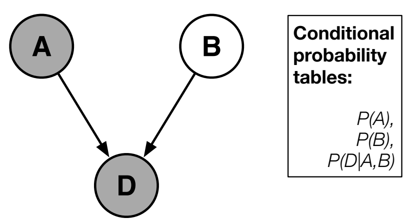
## 3.1 Inference in a polytree

Consider the belief network shown below. In this problem you will be guided through an efficient computation of the posterior probability $P(F|A,C,D,G)$. You are expected to perform these computations *efficiently*—that is, by exploiting the structure of the DAG and not marginalizing over more variables than necessary. Note that the result in each part of this problem will be used to simplify the calculation in the next one. *Justify your steps briefly for full credit.*



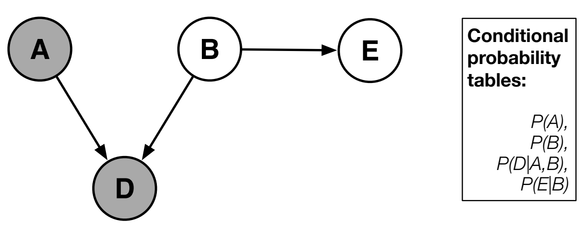

(a) **Bayes rule**

Consider just the part of the belief network shown below. Show how to compute the posterior probability $P(B\!=\!b|A,D)$ in terms of the conditional probability tables (CPTs) shown below.



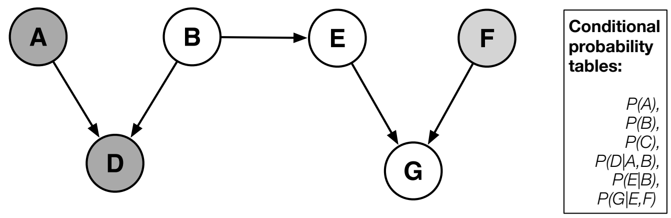

(b) **Marginalization**

Consider just the part of the belief network shown below. Show how to compute the posterior probability $P(E\!=\!e|A,D)$ *in terms of your answer from part (a)* and the CPTs (as needed) of this belief network.



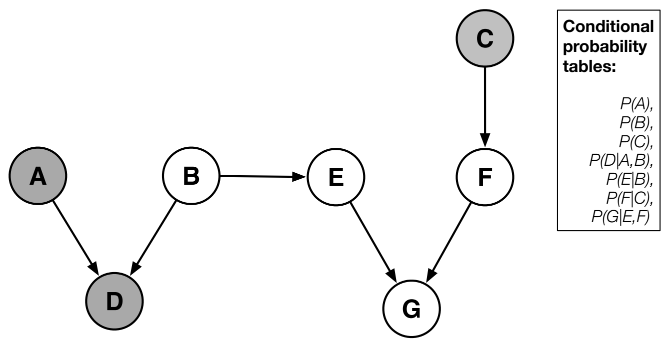
(c) **Marginalization**

Consider the belief network shown below. Show how to compute the posterior probability $P(G|A, D, F\!=\!f)$ *in terms of your answer from part (b)* and the CPTs (as needed) of this belief network.



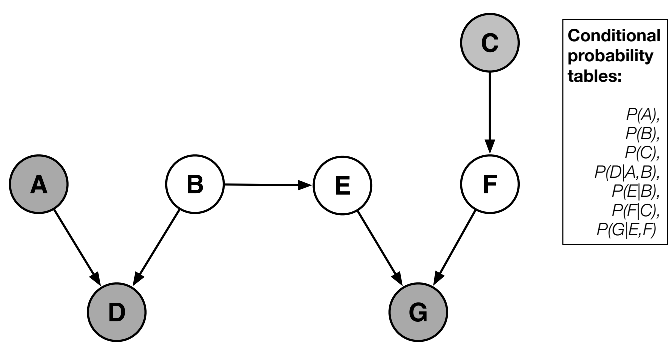

(d) **Marginalization**

Consider the belief network shown below. Show how to compute the posterior probability $P(G|A, C, D)$ *in terms of your answer from part (c)* and the CPTs (as needed) of this belief network.




(e) **Bayes rule**

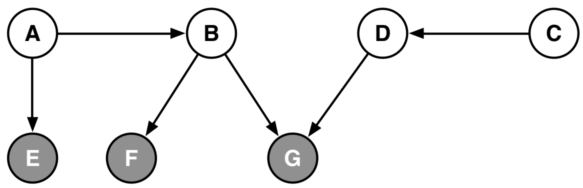
Finally, show how to compute the posterior probability $P(F|A, C, D, G)$ *in terms of your answers from parts (c) and (d)* and the CPTs (as needed) of this belief network.




---

### 3.2 More inference (but with fewer hints)

For the belief network shown below, consider how to *efficiently* compute the posterior probability $P(D|E, F, G)$. This can be done in five consecutive steps in which the later steps rely on the results from earlier ones. In particular, at each step, you'll want to exploit what you just computed in the last one.




Complete the procedure below for this inference; in particular, show how to compute the necessary result, *as efficiently as possible*, at each step. *Show your work and justify your reasoning* for full credit. Your answers should be expressed in terms of the CPTs of the belief network and (as needed) the results of previous steps.

(a) $P(A = a|E)$

(b) $P(B = b|E)$

(c) $P(D = d)$

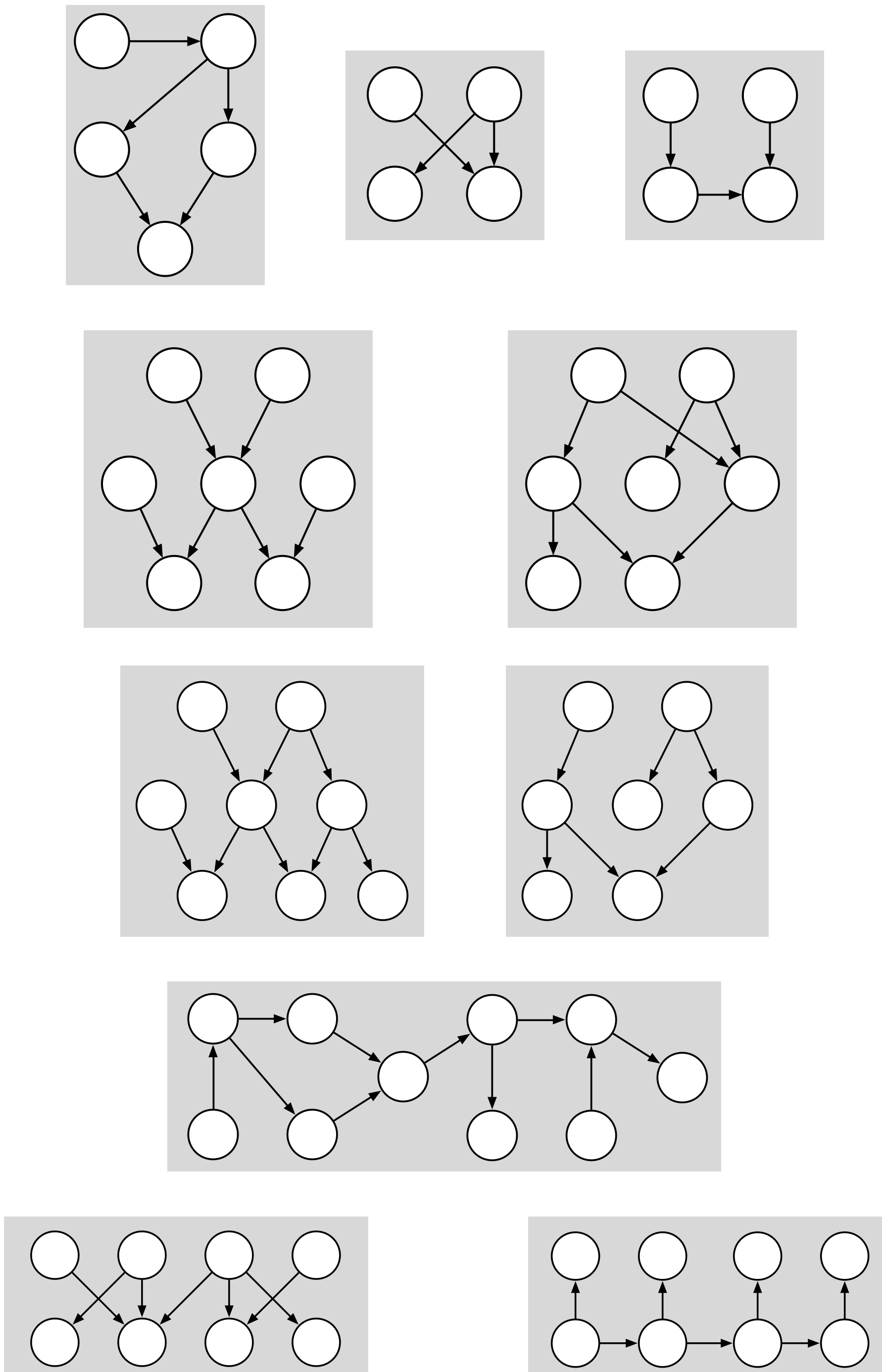
(d) $P(F, G|D=d, E)$

(e) $P(F, G|E)$

(f) $P(D|E, F, G)$

## 3.3 To be, or not to be, a polytree: that is the question.

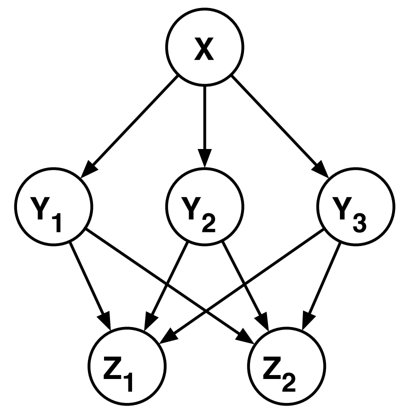
Circle the DAGs shown below that are polytrees. In the other DAGs, shade **two** nodes that could be *clustered* so that the resulting DAG is a polytree.



## 3.4 Node clustering

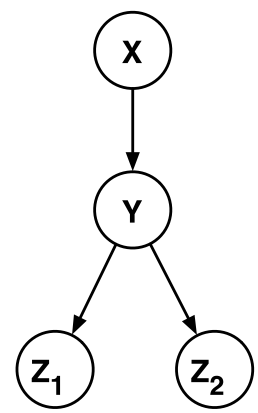
Consider the belief network shown below over binary variables $X$, $Y_1$, $Y_2$, $Y_3$, $Z_1$, and $Z_2$. The network can be transformed into a polytree by clustering the nodes $Y_1$, $Y_2$, and $Y_3$ into a single node $Y$. From the CPTs in the original belief network, fill in the missing elements of the CPTs for the polytree.

| $X$ | $P(Y_1 = 1\|X)$ | $P(Y_2 = 1\|X)$ | $P(Y_3 = 1\|X)$ |
|---|---|---|---|
| 0 | 0.75 | 0.50 | 0.25 |
| 1 | 0.50 | 0.25 | 0.75 |




| $Y_1$ | $Y_2$ | $Y_3$ | $P(Z_1 = 1\|Y_1, Y_2, Y_3)$ | $P(Z_2 = 1\|Y_1, Y_2, Y_3)$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 0.9 | 0.1 |
| 1 | 0 | 0 | 0.8 | 0.2 |
| 0 | 1 | 0 | 0.7 | 0.3 |
| 0 | 0 | 1 | 0.6 | 0.4 |
| 1 | 1 | 0 | 0.5 | 0.5 |
| 1 | 0 | 1 | 0.4 | 0.6 |
| 0 | 1 | 1 | 0.3 | 0.7 |
| 1 | 1 | 1 | 0.2 | 0.8 |

| $Y_1$ | $Y_2$ | $Y_3$ | $Y$ | $P(Y\|X{=}0)$ | $P(Y\|X{=}1)$ | $P(Z_1{=}1\|Y)$ | $P(Z_2{=}1\|Y)$ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | | | | |
| 1 | 0 | 0 | 2 | | | | |
| 0 | 1 | 0 | 3 | | | | |
| 0 | 0 | 1 | 4 | | | | |
| 1 | 1 | 0 | 5 | | | | |
| 1 | 0 | 1 | 6 | | | | |
| 0 | 1 | 1 | 7 | | | | |
| 1 | 1 | 1 | 8 | | | | |



### 3.5 Maximum likelihood estimation for an $n$-sided die

A $n$-sided die is tossed $T$ times, and the results recorded as data. Assume that the tosses $\{x^{(1)}, x^{(2)}, \ldots, x^{(T)}\}$ are identically distributed (i.i.d.) according to the probabilities $p_k = P(X=k)$ of the die, and suppose that over the course of $T$ tosses, the $k^{\text{th}}$ side of the die is observed $C_k$ times.

(a) **Log-likelihood**

Express the log-likelihood $\mathcal{L} = \log P(\text{data})$ of the observed results $\{x^{(1)}, x^{(2)}, \ldots, x^{(T)}\}$ in terms of the probabilities $p_k$ and the counts $C_k$. In particular, show that

$$\mathcal{L}(p) \;=\; \sum_{k=1}^{n} C_k \log p_k.$$

(b) **KL distance**

Define the distribution $q_k = C_k/T$, where $T = \sum_k C_k$ is the total number of counts. Show that

$$\mathrm{KL}(q,p) \;=\; \sum_{k=1}^{n} q_k \log q_k \;-\; \frac{\mathcal{L}(p)}{T},$$

where $\mathrm{KL}(q,p)$ is the KL distance from homework problem 1.6. Conclude that maximizing the log-likelihood $\mathcal{L}$ in terms of the probabilities $p_k$ is equivalent to minimizing the KL distance $\mathrm{KL}(q,p)$ in terms of these same probabilities. (Why is this true?)

(c) **Maximum likelihood estimation**

Recall from homework problem 1.6b that $\mathrm{KL}(q,p) \geq 0$ with equality if and only if $q_k = p_k$ for all $k$. Use this to argue that $p_k = C_k/T$ is the maximum-likelihood estimate—i.e., the distribution that maximizes the log-likelihood of the observed data.

*Note: you have done something quite clever and elegant here, perhaps without realizing it! Generally speaking, to maximize the log-likelihood in part (a), you need to solve the following optimization:*

$$\textbf{Maximize} \quad \sum_k C_k \log p_k \quad \textbf{subject to} \quad \sum_k p_k = 1 \quad \textbf{and} \quad p_k \geq 0.$$

*This is a problem in multivariable calculus requiring the use of Lagrange multipliers to enforce the constraint. You have avoided this calculation by a clever use of the result in homework problem 1.6.*