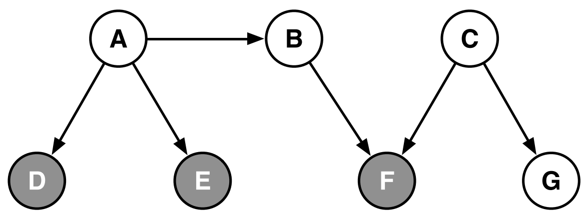
CSE 150. Assignment 5 *Summer 2016*

Out: *Tue Jul 12*
Due: *Fri Jul 15 (outside CSE 3214, by 2 pm)*

5.1 Inference in a polytree

For the belief network shown below, consider how to *efficiently* compute the posterior probability $P(G|D, E, F)$. This can be done in five consecutive steps in which the later steps rely on the results from earlier ones.



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 the results of previous steps.

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

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

(c) $P(F|C = c, D, E)$

(d) $P(C = c|D, E, F)$

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

Hint: *at each step, you'll want to exploit what you just computed in the last one.*

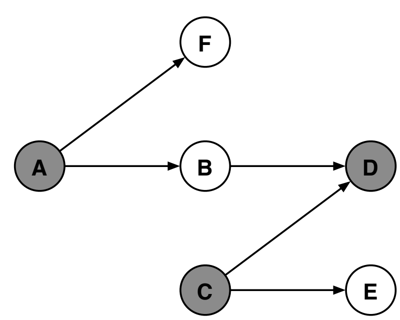
5.2 More inference

For the belief network shown below, consider how to *efficiently* compute the following probabilities. Express your answers in terms of the CPTs of the belief network; also, in parts (b,c), you may re-use your answer from (a). *Justify briefly each step in your calculations.*

(a) $P(B|A, C, D)$

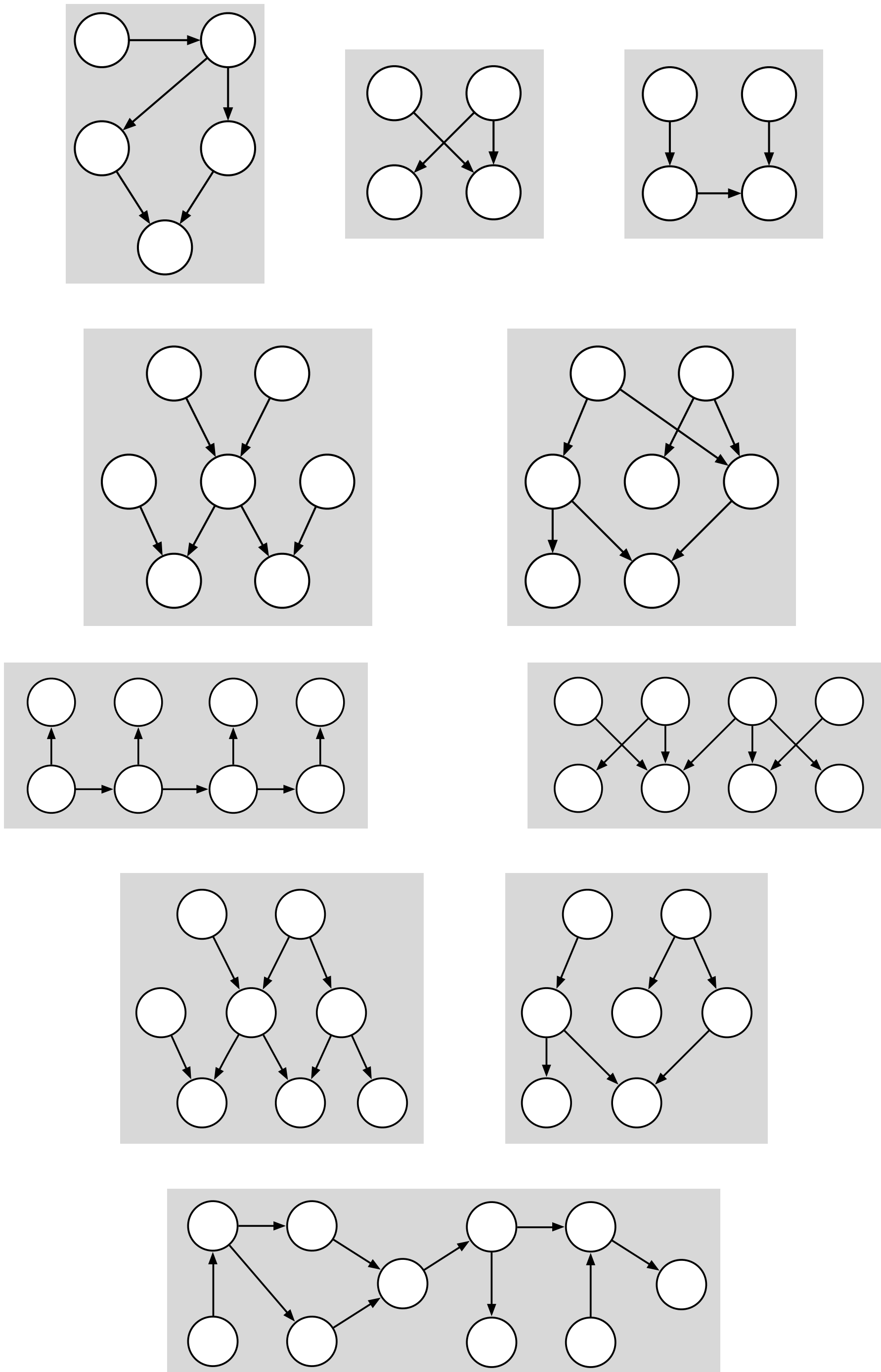
(b) $P(B|A, C, D, E, F)$

(c) $P(B, E, F|A, C, D)$



5.3 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.



5.4 Markov modeling

In this problem, you will construct and compare unigram and bigram models defined over the four-letter alphabet $\mathcal{A} = \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}\}$. Consider the following 16-token sequence $\mathcal{S}$:

$$\mathcal{S} = \text{“}\texttt{a a c c b b d d d d b b c c a a}\text{”}$$

(a) **Unigram model**

Let τ_ℓ denote the ℓth token of this sequence, and let $L = 16$ denote the total sequence length. The overall likelihood of this sequence under a unigram model is given by:

$$P_U(\mathcal{S}) = \prod_{\ell=1}^{L} P_1(\tau_\ell),$$

where $P_1(\tau)$ is the unigram probability for the token $\tau \in \mathcal{A}$. Compute the maximum likelihood estimates of these unigram probabilities on the training sequence $\mathcal{S}$. Complete the table with your answers.

τ	a	b	c	d
$P_1(\tau)$				

(b) **Bigram model**

The overall likelihood of the sequence $\mathcal{S}$ under a bigram model is given by:

$$P_B(\mathcal{S}) = P_1(\tau_1) \prod_{\ell=2}^{L} P_2(\tau_\ell | \tau_{\ell-1}),$$

where $P_2(\tau'|\tau)$ is the bigram probability that token $\tau \in \mathcal{A}$ is followed by token $\tau' \in \mathcal{A}$. Compute the maximum likelihood estimates of these bigram probabilities on the training sequence $\mathcal{S}$. Complete the table with your answers.

τ' (columns); τ (rows)

$P_2(\tau'\|\tau)$	a	b	c	d
a	$\frac{2}{3}$	0	$\frac{1}{3}$	0
b				
c				
d				

(c) **Likelihoods**

Consider again the training sequence $\mathcal{S}$, as well as three test sequences $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$ of the same length, shown below. Note that $\mathcal{T}_2$ and $\mathcal{T}_3$ contain bigrams (<u>underlined</u>) that are not in the training sequence $\mathcal{S}$.

$$
\begin{aligned}
\mathcal{S} &= \text{“}\texttt{a a c c b b d d d d b b c c a a}\text{”} \\
\mathcal{T}_1 &= \text{“}\texttt{b d b d b d b d b d b d b d b d}\text{”} \\
\mathcal{T}_2 &= \text{“}\texttt{a a a }\underline{\texttt{a d}}\texttt{ d d d b b b b c c c c}\text{”} \\
\mathcal{T}_3 &= \text{“}\underline{\texttt{a d a d a d a d a d a d a d a d}}\text{”}
\end{aligned}
$$

Consider the probabilities of these sequences under the unigram and bigram models from parts (a) and (b) of this problem (i.e., the models that you estimated from the training sequence $\mathcal{S}$). For each of the following, indicate whether the probability on the left is equal ($=$), greater ($>$), or less ($<$) than the probability on the right.

<u>*Note: you can (and should) answer these questions without explicitly computing the numerical values of the expressions on the left and right hand sides.*</u>

$$P_U(\mathcal{S}) \quad \square \quad P_U(\mathcal{T}_1)$$

$$P_U(\mathcal{S}) \quad \square \quad P_U(\mathcal{T}_2)$$

$$P_U(\mathcal{S}) \quad \square \quad P_U(\mathcal{T}_3)$$

$$P_B(\mathcal{S}) \quad \square \quad P_B(\mathcal{T}_1)$$

$$P_B(\mathcal{S}) \quad \square \quad P_B(\mathcal{T}_2)$$

$$P_B(\mathcal{T}_2) \quad \square \quad P_B(\mathcal{T}_3)$$

$$P_U(\mathcal{S}) \quad \square \quad P_B(\mathcal{S})$$

$$P_U(\mathcal{T}_1) \quad \square \quad P_B(\mathcal{T}_1)$$

$$P_U(\mathcal{T}_2) \quad \square \quad P_B(\mathcal{T}_2)$$

$$P_U(\mathcal{T}_3) \quad \square \quad P_B(\mathcal{T}_3)$$

(d) **Likelihoods**

Consider the model obtained by linear interpolation (or mixing) of the unigram and bigram models estimated in part (a) of this problem:

$$P_M(\tau'|\tau) \;=\; (1-\lambda)P_1(\tau') + \lambda P_2(\tau'|\tau),$$

with mixing coefficient $\lambda \in [0,1]$. For a sequence of tokens of length L, the mixture model computes the log-likelihood as:

$$\mathcal{L} \;=\; \log P_1(\tau_1) + \sum_{\ell=2}^{L} \log P_M(\tau_\ell|\tau_{\ell-1}).$$

Naturally, this value varies as a function of the coefficient λ. For λ near zero, it is close to the log-likelihood of the unigram model; for λ near one, it is close to that of the bigram model. This last part of this problem asks you to consider, for each of the sequences below, the *qualitative* behavior of the mixture model's log-likelihood as a function of $\lambda \in [0,1]$. (For instance, is this function constant, or if not, where do its maximum and minimum occur?)

The plots below illustrate four possible behaviors of the mixture model's log-likelihood as a function of $\lambda \in [0,1]$. For each sequence below, indicate the one plot (either A, B, C, or D) that sketches the correct qualitative behavior.

$\mathcal{S}$ = "a a c c b b d d d d b b c c a a" ☐

$\mathcal{T}_1$ = "b d b d b d b d b d b d b d b d" ☐

$\mathcal{T}_2$ = "a a a $\underline{\text{a d}}$ d d d b b b b c c c c" ☐

$\mathcal{T}_3$ = "$\underline{\text{a d a d a d a d a d a d a d a d}}$" ☐