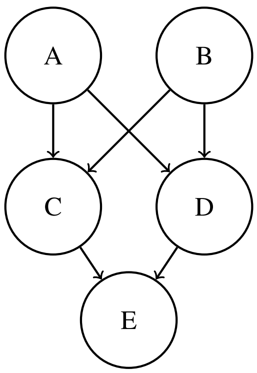
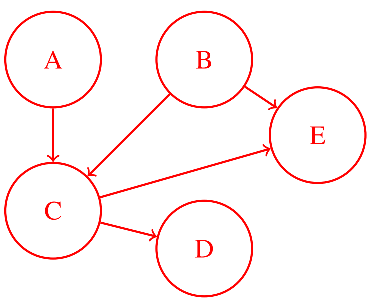
Q1. Bayes Nets and Joint Distributions

(a) Write down the joint probability distribution associated with the following Bayes Net. Express the answer as a product of terms representing individual conditional probabilities tables associated with this Bayes Net:



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

(b) Draw the Bayes net associated with the following joint distribution:
$P(A) \cdot P(B) \cdot P(C|A,B) \cdot P(D|C) \cdot P(E|B,C)$



(c) Do the following products of factors correspond to a valid joint distribution over the variables A, B, C, D? (Circle FALSE or TRUE.)

(i) **[FALSE]** TRUE $P(A) \cdot P(B) \cdot P(C|A) \cdot P(C|B) \cdot P(D|C)$

(ii) FALSE **[TRUE]** $P(A) \cdot P(B|A) \cdot P(C) \cdot P(D|B,C)$

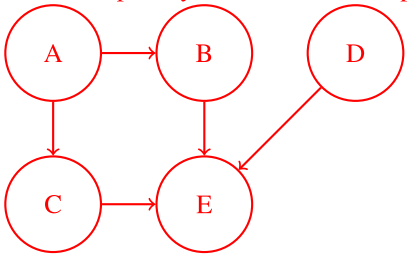
(iii) **[FALSE]** TRUE $P(A) \cdot P(B|A) \cdot P(C) \cdot P(C|A) \cdot P(D)$

(iv) **[FALSE]** TRUE $P(A|B) \cdot P(B|C) \cdot P(C|D) \cdot P(D|A)$

(d) What factor can be multiplied with the following factors to form a valid joint distribution? (Write "none" if the given set of factors can't be turned into a joint by the inclusion of exactly one more factor.)

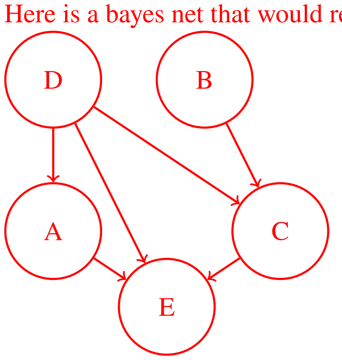
(i) $P(A) \cdot P(B|A) \cdot P(C|A) \cdot P(E|B, C, D)$

P(D) is missing. D could also be conditioned on A,B, and/or C without creating a cycle (e.g. $P(D|A, B, C)$). Here is an example bayes net that would represent the distribution after adding in $P(D)$:



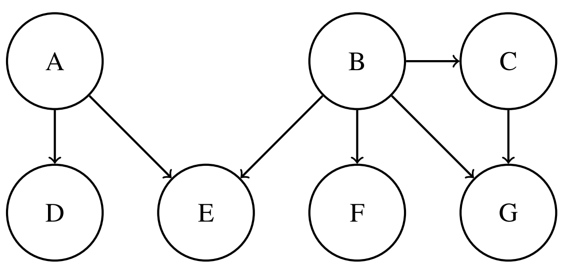
(ii) $P(D) \cdot P(B) \cdot P(C|D, B) \cdot P(E|C, D, A)$

P(A) is missing to form a valid joint distributions. A could also be conditioned on B, C, and/or D (e.g. $P(A|B, C, D)$). Here is a bayes net that would represent the distribution is $P(A|D)$ was added in.



(e) Answer the next questions based off of the Bayes Net below:

All variables have domains of {-1, 0, 1}



(i) Before eliminating any variables or including any evidence, how many entries does the factor at G have?

The factor is $P(G|B, C)$, so that gives $3^3 = 27$ entries.

(ii) Now we observe $e = 1$ and want to query $P(D|e = 1)$, and you get to pick the first variable to be eliminated.

- Which choice would create the **largest** factor f_1?

  Eliminating B first would give the largest f_1:, $f_1(A, F, G, C, e) = \sum_{B=b} P(b)P(e|A, b)P(F|b)P(G|b, C)P(C|b)$. This factor has 3^4 entries.

- Which choice would create the **smallest** factor f_1?

  eliminating F first would give smallest factors of 3 entries: $f_1(B) = \sum_f P(f|B)$. Eliminating D is not correct because D is the query variable.

Q2. Probability and Bayes Nets

(a) A, B, and C are random variables with binary domains. How many entries are in the following probability tables and what is the sum of the values in each table? Write a "?" in the box if there is not enough information given.

Table	Size	Sum
$P(A, B \mid C)$	8	2
$P(A \mid +b, +c)$	2	1
$P(+a \mid B)$	2	?

(b) Circle true if the following probability equalities are valid and circle false if they are invalid (leave it blank if you don't wish to risk a guess). Each True/False question is worth 1 points. Leaving a question blank is worth 0 points. **Answering incorrectly is worth −1 points.**

No independence assumptions are made.

(i) [*true* or *false*] $P(A, B) = P(A|B)P(A)$
False. $P(A, B) = P(A|B)P(B)$ would be a valid example.

(ii) [*true* or *false*] $P(A|B)P(C|B) = P(A, C|B)$
False. This assumes that A and C are conditionally independent given B.

(iii) [*true* or *false*] $P(B, C) = \sum_{a \in A} P(B, C|A)$
False. $P(B, C) = \sum_{a \in A} P(A, B, C)$ would be a valid example.

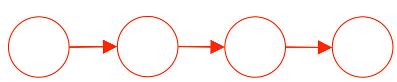
(iv) [*true* or *false*] $P(A, B, C, D) = P(C)P(D|C)P(A|C, D)P(B|A, C, D)$
True. This is a valid application of the chain rule.

(c) Space Complexity of Bayes Nets

Consider a joint distribution over N variables. Let k be the domain size for all of these variables, and let d be the maximum indegree of any node in a Bayes net that encodes this distribution.

(i) What is the space complexity of storing the entire joint distribution? Give an answer of the form $O(\cdot)$.
$O(k^N)$ was the intended answer. Because of the potentially misleading wording, we also allowed $O(N k^{d+1})$, one possible bound on the space complexity of storing the Bayes net ($O((N-d)k^{d+1})$ is an asymptotically tighter bound, but this requires considerably more effort to prove).

(ii) Draw an example of a Bayes net over four binary variables such that it takes less space to store the Bayes net than to store the joint distribution.
A simple Markov chain works. Size $2+4+4+4 = 14$, which is less than $2^4 = 16$. Less edges, less inbound edges (v-shape), or no edges would work too.



(iii) Draw an example of a Bayes net over four binary variables such that it takes more space to store the Bayes net than to store the joint distribution.
Size $2+2+2+2^4 = 22$, which is more than $2^4 = 16$. Other configurations could work too, especially any with a node with indegree 3.