Stanford University

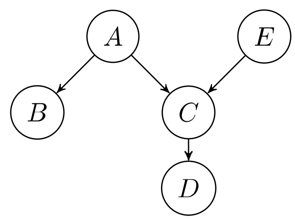
AA228/CS238: Decision Making Under Uncertainty
Fall 2021

Prof. Mykel J. Kochenderfer • Online • email: *mykel@stanford.edu*

QUIZ 1 **Due date: October 8, 2021 (5pm Pacific)**

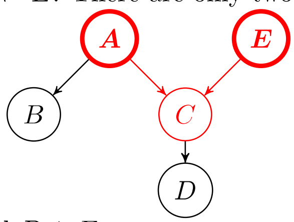
Quizzes will be taken on Gradescope. You may consult any material (e.g., books, calculators, computer programs, and online resources), but you may not consult other people inside or outside of the class. The quiz is designed to be completed in 60 minutes, but we will grant you 90 minutes total to complete and submit your quiz (including uploading any images, handling any logistical issues, etc.) The timing on Gradescope is a hard cutoff. You can start at 5pm PDT on Thursday. To accommodate those in other timezones and complex working situations, the quizzes will be open until 5pm PDT on Friday. Ed will not allow any public posts during that time. Partial credit may be awarded based on work shown. **Out of fairness to all students, only material submitted during the allowed time will be graded.**

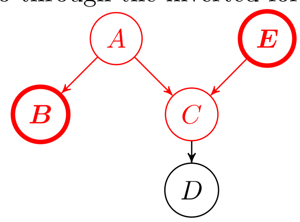
Question 1. Suppose we have a Bayesian network with the following structure:



Identify all pairs of variables that are independent of each other (given no evidence).

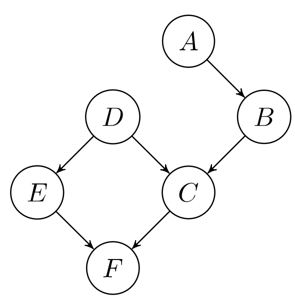
Solution: The only d-separation rule that would lead to independence without evidence requires that an inverted fork must be part of the path. Due to the absence of evidence, it is implied that neither the child node or its descendants could be in the set of evidence nodes. There is one inverted fork in the Bayesian network: $A \to C \leftarrow E$. There are only two paths in the network that go through the inverted fork:





Hence, $A \perp E$ and $B \perp E$.

Question 2. Suppose we have a Bayesian network with the following structure:



We begin to apply the sum-product variable elimination algorithm using *reverse* alphabetic ordering to infer $P(b^1 \mid e^0)$. Show the expression for computing the first new factor produced when eliminating F.

Solution: The following factors can be found from the Bayesian network: $\phi_1(A)$, $\phi_2(A, B)$, $\phi_3(D)$, $\phi_4(B, C, D)$, $\phi_5(D, E)$, and $\phi_6(C, E, F)$. Evidence is at E for this problem. Using this information, we can find the factors $\phi_7(D)$ and $\phi_8(C, F)$. For the elimination of F, we find the desired factor $\phi_9(C)$ by summing over ϕ_8: $\phi_9(C) = \sum_f \phi_8(f, C)$.

Question 3. Suppose we developed a new rocket and we are unsure about its failure probability. Prior to our testing campaign, we start with a uniform prior over the failure probability. We then launch k rockets without observing any failures. What must k be in order for us to assign exactly 20% probability to the next rocket failing?

Solution: We model the posterior distribution over failure probability using a Beta distribution. The uniform prior would be represented by a Beta(1, 1) distribution. If we observe k launches without failure, then the posterior distribution over the failure probability is Beta$(1, 1 + k)$. The probability that the next rocket fails can be computed using the law of total probability:

$$P(\text{fail} \mid k \text{ successes}) = \int P(\text{fail} \mid \theta) P(\theta \mid k \text{ successes}) d\theta = \int \theta \cdot \text{Beta}(\theta \mid 1, 1 + k) d\theta$$

This probability corresponds to the expectation of the Beta$(1, 1 + k)$ distribution which is:

$$\frac{1}{1 + 1 + k} = \frac{1}{2 + k}$$

Setting this expectation to 0.2 and solving for k, we get $k = 3$.

Question 4. Suppose we want to compute the Bayesian score for the network $A \rightarrow B$ with binary variables ignoring the contribution of $\log P(G)$. We have a small dataset consisting of $\{(a^1, b^1), (a^1, b^0)\}$. Write down an expression for this score in terms of $g_1, g_2, \ldots$, where $g_i = \log \Gamma(i)$. Assume a uniform Dirichlet prior over parameters as done in the K2 algorithm.

Solution: The data matrix D (samples represent columns) and the count matrices M_A and M_B are:

$$D = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \qquad M_A = \begin{bmatrix} 0 & 2 \end{bmatrix} \qquad M_B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

Since the prior is uniform, we have $\alpha_{ijk} = 1$ for all i, j, and k. Calculating α_{ij0} and m_{ij0} leads to:

$$\alpha_{A10} = \sum_{k=1}^{2} \alpha_{A1k} = 2 \qquad \alpha_{B10} = \alpha_{B20} = \sum_{k=1}^{2} \alpha_{B1k} = \sum_{k=1}^{2} \alpha_{B2k} = 2$$

$$m_{A10} = \sum_{k=1}^{2} m_{A1k} = 2 \qquad m_{B10} = \sum_{k=1}^{2} m_{B1k} = 0 \qquad m_{B20} = \sum_{k=1}^{2} m_{B2k} = 2$$

The Bayesian score then is:

$$\log(P(G \mid D)) = \log(1) + \underbrace{\log\left(\frac{\Gamma(\alpha_{A10})}{\Gamma(\alpha_{A10} + m_{A10})}\right) + \log\left(\frac{\Gamma(\alpha_{A11} + m_{A11})}{\Gamma(\alpha_{A11})}\right) + \log\left(\frac{\Gamma(\alpha_{A12} + m_{A12})}{\Gamma(\alpha_{A12})}\right)}_{\text{Node } A}$$
$$+ \underbrace{\log\left(\frac{\Gamma(\alpha_{B10})}{\Gamma(\alpha_{B10} + m_{B10})}\right) + \log\left(\frac{\Gamma(\alpha_{B11} + m_{B11})}{\Gamma(\alpha_{B11})}\right) + \log\left(\frac{\Gamma(\alpha_{B12} + m_{B12})}{\Gamma(\alpha_{B12})}\right)}_{\text{Node B with } j=1}$$
$$+ \underbrace{\log\left(\frac{\Gamma(\alpha_{B20})}{\Gamma(\alpha_{B20} + m_{B20})}\right) + \log\left(\frac{\Gamma(\alpha_{B21} + m_{B21})}{\Gamma(\alpha_{B21})}\right) + \log\left(\frac{\Gamma(\alpha_{B22} + m_{B22})}{\Gamma(\alpha_{B22})}\right)}_{\text{Node } B \text{ with } j=2}$$

Using the values that we calculated leads to:

$$
\begin{aligned}
\log(P(G \mid D)) = 0 &+ \log\left(\frac{\Gamma(2)}{\Gamma(2+2)}\right) + \underbrace{\log\left(\frac{\Gamma(1+0)}{\Gamma(1)}\right)}_{=0} + \log\left(\frac{\Gamma(1+2)}{\Gamma(1)}\right) \\
&+ \underbrace{\log\left(\frac{\Gamma(2)}{\Gamma(2+0)}\right)}_{=0} + \underbrace{\log\left(\frac{\Gamma(1+0)}{\Gamma(1)}\right)}_{=0} + \underbrace{\log\left(\frac{\Gamma(1+0)}{\Gamma(1)}\right)}_{=0} \\
&+ \log\left(\frac{\Gamma(2)}{\Gamma(2+2)}\right) + \log\left(\frac{\Gamma(1+1)}{\Gamma(1)}\right) + \log\left(\frac{\Gamma(1+1)}{\Gamma(1)}\right)
\end{aligned}
$$

This can be rewritten as:

$$
\begin{aligned}
\log(P(G \mid D)) &= \log(\Gamma(2)) - \log(\Gamma(4)) + \log(\Gamma(3)) - \log(\Gamma(1)) + \log(\Gamma(2)) \\
&\quad - \log(\Gamma(4)) + \log(\Gamma(2)) - \log(\Gamma(1)) + \log(\Gamma(2)) - \log(\Gamma(1))
\end{aligned}
$$

Finally, using the abbreviation $g_i = \log(\Gamma(i))$, this leads to:

$$
\log(P(G \mid D)) = -2g_4 + g_3 + \underbrace{4g_2}_{=0} - \underbrace{3g_1}_{=0} = -2g_4 + g_3.
$$

Question 5. Suppose someone is infected with a disease. They have a 10% chance of recovery each day. What is the expected number of days until recovery? Write an equation (including summations is okay) to compute this value. You do not actually have to compute this value.

Solution: We use s to represent sick and r to represent recovered. The probability of being sick the next day given that we are currently sick is $P(s \mid s) = 0.9$. The probability of recovering the next day given that we are currently sick is $P(r \mid s) = 0.1$. The probability of recovering in exactly k days is $P(r \mid s)P(s \mid s)^{k-1}$ (with $k-1$ days sick and recovering on the kth day). The expected value of the number of days until recovery is then

$$
\sum_{k=1}^{\infty} kp(r \mid s)p(s \mid s)^{k-1} = \sum_{k=1}^{\infty} 0.1k \cdot 0.9^{k-1}
$$

(Not required for full credit: It turns out that this is equivalent to computing the mean of a geometric distribution with parameter $p = 0.1$. The mean of a geometric distribution is $1/p$, which in this case is 10.)