Homework 1

CS 6347: Statistical methods in AI/ML

Instructor: Vibhav Gogate

Vibhav.Gogate@utdallas.edu

Student: Xiaodi Li

Net ID: XXL170011

[Key: AD: Book by Adnan Darwiche, KF: Koller Friedman]

- **WARNING:** Start early. Some problems are quite hard (e.g., Problem 10).
- **Each problem is worth 10 points**

Problem 1: Propositional logic [Problems 2.6 to 2.9 from AD]

- Prove the Refutation theorem, namely $\alpha \vDash \beta$ iff $\alpha \wedge \neg\beta$ is inconsistent.
  Proof:
  (1) $\alpha \vDash \beta$ means that sentence β is true at world α. Then, $\neg\beta$ is false at world α. Thus, $\alpha \wedge \neg\beta$ is also false, which means $\alpha \wedge \neg\beta$ is inconsistent.
  (2) If $\alpha \wedge \neg\beta$ is inconsistent, $\neg\beta$ is false at world α. Then, β is true at world α. Thus, $\alpha \vDash \beta$.

  Proof is done.

- Prove the Deduction theorem, namely $\alpha \vDash \beta$ iff $\alpha \Rightarrow \beta$ is valid.
  Proof:
  (1) $\alpha \vDash \beta$ means that sentence β is true at world α. Thus, $\alpha \Rightarrow \beta$ is true in all models, which means it is valid.
  (2) If $\alpha \Rightarrow \beta$ is valid, β is true at world α. Thus, $\alpha \vDash \beta$.

  Proof is done.

- Prove that if $\alpha \vDash \beta$ then $\alpha \wedge \beta$ is equivalent to α.
  Proof:
  If $\alpha \vDash \beta$, then sentence β is true at world α. Thus, $\alpha \wedge \beta$ is equivalent to $\alpha \wedge true$, which is equivalent to α.
  Proof done.
- Prove that if $\alpha \vDash \beta$ then $\alpha \vee \beta$ is equivalent to β.
  Proof:
  If $\alpha \vDash \beta$, then sentence β is true at world α. Thus, $\alpha \vee \beta$ is equivalent to $\alpha \vee true$, which is equivalent to $true$, which is equivalent to β.
  Proof is done.

Problem 2: Probability theory

- Prove that the following two definitions of conditional independence are equivalent.
    1. $Pr(\alpha|\beta \wedge \gamma) = Pr(\alpha|\gamma)$
    2. $Pr(\alpha \wedge \beta|\gamma) = Pr(\alpha|\gamma)\, Pr(\beta|\gamma)$

  [**Hint:** Derive one using the other using laws of probability.]

  Proof:

  (1) According to the definition of conditional independence, $Pr(\alpha|\beta \wedge \gamma) = Pr(\alpha|\gamma)$ means that α and β are conditionally independent given γ.

  (2) According to the definition of independence of two events, $Pr(\alpha \wedge \beta|\gamma) = Pr(\alpha|\gamma)\, Pr(\beta|\gamma)$ means that $\alpha|\gamma$ and $\beta|\gamma$ are independent, which also means that α and β are conditionally independent given γ.

  Proof is done.

Problem 3: Probability theory

We solved the following problem in class.

- **Fact 1:** The probability that your partner fails a lie detector test given that he/she is cheating on you is 0.98 (or 98%). The probability that your partner fails the test given that he/she is not cheating on you is 0.02.
- **Fact 2:** You are a CS graduate student.
- **Fact 3:** You should break up with your partner if he/she is cheating on you with a probability greater than 0.05.

Today, you found that you failed the lie detector test. You are in panic mode and are sure that your partner will break up with you. Suddenly you realize that you had previously found the following, pretty reliable statistic on the internet: only 1 out of 10000 CS graduate students (since they are boring people) cheat on their partners. Given this new information and using sound probabilistic arguments, how do you convince your partner that you are not cheating on him/her?

We derived in class that with this new information the probability of cheating given that you have failed the test is much smaller than 0.05.

Here is the twist; your partner is not totally convinced. He/She tells you to take the test three more times on three different days. Unfortunately, you fail two out of the three tests. Can you still convince your partner that you are not cheating on him/her (assume that the three tests are independent of each other). Justify your answer using purely probabilistic arguments.

Answer:

Notation: T: Detector test, c: cheating.

According to Fact 1, $P(T = fail|c = yes) = 0.98$. According to Fact 2, $P(T = fail|c = no) = 0.02$. Moreover, $P(c = yes) = 1/10000$.

We need to compute $P(c = yes|T_1 = fail, T_2 = fail, T_3 = success) = C_3^2 P(c = yes|T = fail) * P(c = yes|T = fail) * P(c = yes|T = success)$

Here,

$$
\begin{aligned}
P(c = yes|T = fail) &= \frac{P(T = fail|c = yes)P(c = yes)}{P(T = fail)} \\
&= \frac{P(T = fail|c = yes)P(c = yes)}{P(T = fail \wedge c = yes) + P(T = fail \wedge c = no)} \\
&= \frac{P(T = fail|c = yes)P(c = yes)}{P(T = fail|c = yes) * P(c = yes) + P(T = fail|c = no) * P(c = no)} \\
&= \frac{0.98 * \frac{1}{10000}}{0.98 * \frac{1}{10000} + 0.02 * \left(1 - \frac{1}{10000}\right)} \\
&\approx 0.0049
\end{aligned}
$$

$$
\begin{aligned}
P(c = yes|T = success) &= \frac{P(T = success|c = yes)P(c = yes)}{P(T = success)} \\
&= \frac{\left(1 - P(T = fail|c = yes)\right) * P(c = yes)}{P(T = success \wedge c = yes) + P(T = fail \wedge c = no)} \\
&= \frac{\left(1 - P(T = fail|c = yes)\right) * P(c = yes)}{P(T = success|c = yes)P(c = yes) + P(T = fail|c = no)P(c = no)} \\
&= \frac{(1 - 0.98) * \frac{1}{10000}}{(1 - 0.98) * \frac{1}{10000} + 0.02 * \left(1 - \frac{1}{10000}\right)} \\
&= 0.0001
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&P(c = yes|T_1 = fail, T_2 = fail, T_3 = success) \\
&= C_3^2 P(c = yes|T = fail) * P(c = yes|T = fail) * P(c = yes|T = success) \\
&= 3 * 0.0049 * 0.0049 * 0.0001 \\
&= 7.203 * 10^{-9} \\
&< 0.05
\end{aligned}
$$

Thus, I can still convince my partner that I am not cheating on him/her.

Problem 4: Probability theory (Exercise 2.10 from Koller & Friedman)

The question investigates the way in which conditional independence relationships affect the amount of information needed for probabilistic calculations. Let α, β, and γ be three propositional variables.

- Suppose we wish to calculate $Pr(\alpha|\beta, \gamma)$ and we have no conditional independence information. Which of the following sets of numbers is sufficient for the calculation?
    1. $Pr(\alpha, \beta)$, $Pr(\alpha)$, $Pr(\beta|\alpha)$ and $Pr(\gamma|\alpha)$.
    2. $Pr(\beta, \gamma)$, $Pr(\alpha)$ and $Pr(\beta, \gamma|\alpha)$
    3. $Pr(\beta|\alpha)$, $Pr(\gamma|\alpha)$ and $Pr(\alpha)$.

For each case, justify your response either by showing how to calculate the desired answer or by explaining why this is not possible.

- Suppose we know that β and γ are conditionally independent given α. Now which of the preceeding three sets is sufficient. Justify your response as before.

Answer:

- $Pr(\alpha|\beta,\gamma) = \frac{\Pr(\alpha,\beta,\gamma)}{\Pr(\beta,\gamma)}$. (1)
  Thus, in order to calculate $Pr(\alpha|\beta,\gamma)$, we need to calculate $\Pr(\alpha,\beta,\gamma)$ and $\Pr(\beta,\gamma)$.
  1. No. Because we have no conditional independent information, we can only calculate $\Pr(\alpha,\beta)$ and $\Pr(\alpha,\gamma)$ based on Bayes rule. This cannot give us $\Pr(\alpha,\beta,\gamma)$ and $\Pr(\beta,\gamma)$. Thus, we cannot calculate these two probabilities given the probabilities we have.
  2. Yes. According to Bayes rule, $\Pr(\alpha,\beta,\gamma) = \Pr(\beta,\gamma|\alpha) * \Pr(\alpha)$, we also know $\Pr(\beta,\gamma)$. Thus, we can calculate $Pr(\alpha|\beta,\gamma)$ based on formula (1).
  3. No. Because we have no conditional independent information, we cannot calculate these two probabilities given the probabilities we have.
- If β and γ are conditionally independent given α, we can know that
  $Pr(\beta|\alpha,\gamma) = Pr(\beta|\alpha)$, $Pr(\gamma|\alpha,\beta) = Pr(\gamma|\alpha)$ (2)
  $\Pr(\beta,\gamma|\alpha) = \Pr(\beta|\alpha) * \Pr(\gamma|\alpha)$ (3)
  1. No. Given $Pr(\beta|\alpha)$ and $Pr(\gamma|\alpha)$, according to formula (3), we can calculate $\Pr(\beta,\gamma|\alpha)$. Then, given $\Pr(\alpha)$, based on Bayes rule, $\Pr(\alpha,\beta,\gamma) = \Pr(\beta,\gamma|\alpha) * \Pr(\alpha)$. However, we cannot calculate $\Pr(\beta,\gamma)$. Thus, we cannot calculate $Pr(\alpha|\beta,\gamma)$.
  2. Yes. According to Bayes rule, $\Pr(\alpha,\beta,\gamma) = \Pr(\beta,\gamma|\alpha) * \Pr(\alpha)$, we also know $\Pr(\beta,\gamma)$. Thus, we can calculate $Pr(\alpha|\beta,\gamma)$ based on formula (1).
  3. No. Given $Pr(\beta|\alpha)$ and $Pr(\gamma|\alpha)$, according to formula (3), we can calculate $\Pr(\beta,\gamma|\alpha)$. Then, given $\Pr(\alpha)$, based on Bayes rule, $\Pr(\alpha,\beta,\gamma) = \Pr(\beta,\gamma|\alpha) * \Pr(\alpha)$. However, we cannot calculate $\Pr(\beta,\gamma)$. Thus, we cannot calculate $Pr(\alpha|\beta,\gamma)$.

Problem 5: Independence relations

- Prove that Weak Union and Contraction hold for any probability distribution Pr.
- Provide a counter-example to the intersection property. (You cannot use the counter example given in AD, you have to make your own.)

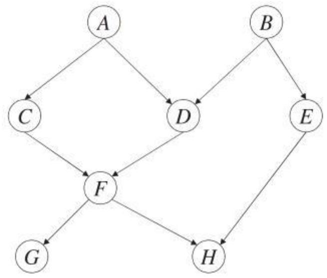
Answer:

- Weak Union: $(X \perp Y,W|Z) \Rightarrow (X \perp Y|Z,W)$.
  Proof:
  According to the definition of conditional independence, $(X \perp Y,W|Z)$ means that $\Pr(X|(Y,W),Z) = \Pr(X|Z)$.
  According to decomposition property, $(X \perp Y,W|Z) \Rightarrow (X \perp Y|Z)$ & $(X \perp W|Z)$.

$(X \perp W|Z)$ means that $\Pr(X|Z, W) = \Pr(X|Z)$.
According to the definition of conditional independence, $\Pr(X|Y, Z, W) = \Pr(X|Z, W)$, which means $\Pr(X|Y, (Z, W)) = \Pr(X|Z, W)$. Thus, $(X \perp Y|Z, W)$.

- Contraction: $(X \perp W|Z, Y)$ & $(X \perp Y|Z) \Rightarrow (X \perp Y, W|Z)$.
  Proof:
  According to the definition of conditional independence, $(X \perp W|Z, Y)$ means that $\Pr(X|W, Z, Y) = \Pr(X|Z, Y)$. $(X \perp Y|Z)$ means that $\Pr(X|Y, Z) = \Pr(X|Z)$. Based on the two equality, we can derive that $\Pr(X|W, Z, Y) = \Pr(X|Z, Y) = \Pr(X|Z)$. Thus, $(X \perp Y, W|Z)$.

Problem 6: Bayesian networks (AD Exercise 4.1)



A	Θ_A
1	.2
0	.8

B	Θ_B
1	.7
0	.3

B	E	$\Theta_{E\mid B}$
1	1	.1
1	0	.9
0	1	.9
0	0	.1

A	B	D	$\Theta_{D\mid AB}$
1	1	1	.5
1	1	0	.5
1	0	1	.6
1	0	0	.4
0	1	1	.1
0	1	0	.9
0	0	1	.8
0	0	0	.2

Consider the Bayesian network given above.

1. List the Markovian assumptions asserted by the DAG.
   (1) $I(A, \emptyset, \{B, E\})$
   (2) $I(B, \emptyset, \{A, C\})$
   (3) $I(C, A, \{B, D, E\})$
   (4) $I(D, \{A, B\}, \{C, E\})$
   (5) $I(E, B, \{A, C, D, F, G\})$
   (6) $I(F, \{C, D\}, \{A, B, E\})$
   (7) $I(G, F, \{A, B, C, D, E\})$
   (8) $I(H, \{E, F\}, \{A, B, C, D, G\})$
2. Express $Pr(a, b, c, d, e, f, g, h)$ in terms of network parameters.
   $\Pr(a, b, c, d, e, f, g, h) = \theta(a)\theta(b)\theta(c|a)\theta(d|b)\theta(e|b)\theta(f|c, d)\theta(g|f)\theta(h|e, f)$
3. Compute $Pr(A = 0, B = 0)$ and $Pr(E = 1|A = 1)$. Justify your answers.
   (1) Because $I(A, \emptyset, \{B, E\})$, $\Pr(A|B, E) = \Pr(A)$. Based on decomposition property, $\Pr(A|B) = \Pr(A)$. Because $I(B, \emptyset, \{A, C\})$, $\Pr(B|A, C) = \Pr(B)$. Based on decomposition property, $\Pr(B|A) = \Pr(B)$. Thus, A and B are

independent of each other. Thus, $\Pr(A=0, B=0) = \Pr(A=0) * \Pr(B=0) = 0.8 * 0.3 = 0.24$.

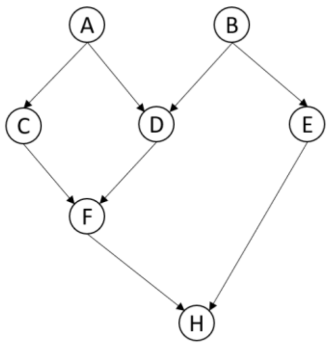
(2) Because $I(A, \emptyset, \{B, E\})$, $\Pr(A|B, E) = \Pr(A)$. Based on decomposition property, $\Pr(A|E) = \Pr(A)$. Thus, A and E are independent of each other. Thus, $\Pr(E=1|A=1) = \Pr(\mathrm{E}=1)$. $\Pr(E=1) = \Pr(E=1, B=0) + \Pr(E=1, B=1) = \Pr(E=1|B=0) * \Pr(B=0) + \Pr(E=1|B=1) * \Pr(B=1) = 0.9 * 0.3 + 0.1 * 0.7 = 0.34$

4. True or false? Why?

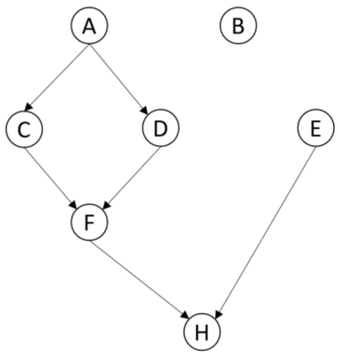
(a) $dsep(A, BH, E)$

True.

First, we delete any leaf node W from DAG G as long as W does not belong to $A \cup E \cup BH$. Then, we will get the following new graph:



Second, we delete all edges outgoing from nodes in BH. Then, we will get the following new DAG G':

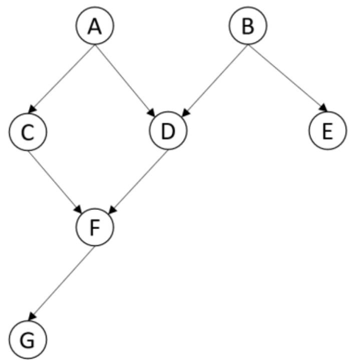


There is no path from A to E in G'. Thus, $dsep(A, BH, E)$ is true.

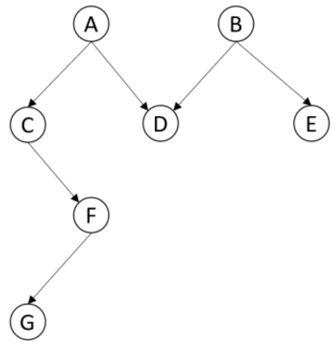
(b) $dsep(G, D, E)$

True.

First, we delete any leaf node W from DAG G as long as W does not belong to $G \cup D \cup E$. Then, we will get the following new graph:



Second, we delete all edges outgoing from nodes in D. Then, we will get the following new DAG G':

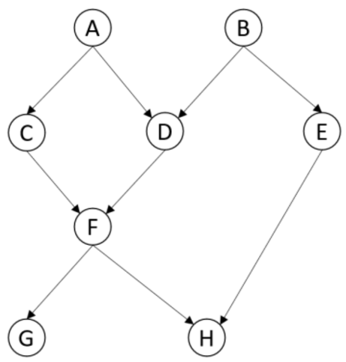


There is no path from G to E in G'. Thus, $dsep(G, D, E)$ is true.

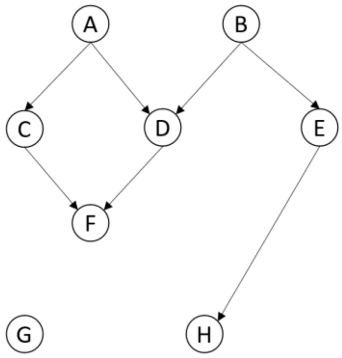
(c) $dsep(AB, F, GH)$

False.

First, we delete any leaf node W from DAG G as long as W does not belong to $AB \cup F \cup GH$. Then, we will get the following graph (there is no change of the graph):



Second, we delete all edges outgoing from nodes in F. Then, we will get the following new DAG G':

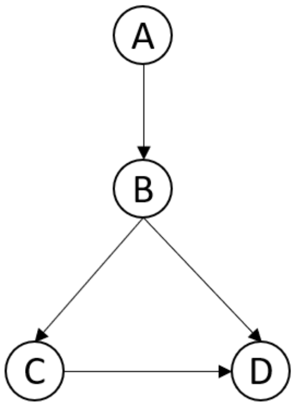


There is still a path $(B \to E \to H)$ from AB to GH. Thus, $dsep(AB, F, GH)$ is false.

Problem 7: Bayesian networks (AD Exercise 4.12)

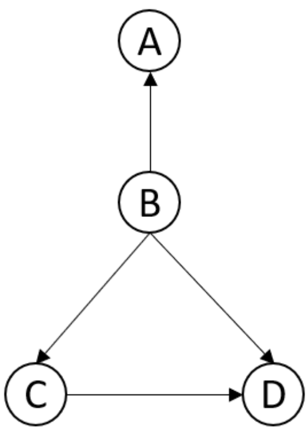
Construct two distinct DAGs over variables A, B, C, and D. Each DAG must have exactly four edges and the DAGs must agree on d-separation.

DAG G_1:



G_1 satisfies $dsep(A, B, CD)$.

DAG G_2:

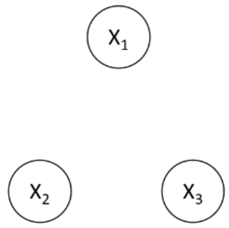


G_2 satisfies $dsep(A, B, CD)$.

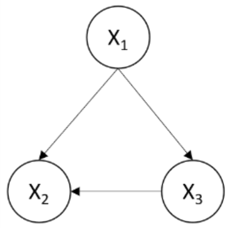
Problem 8: Bayesian networks (AD Exercise 4.15)

Identify a DAG that is a D-MAP for all distributions Pr over variables X. Similarly, identify another DAG that is an I-MAP for all distributions Pr over variables X.

1. If a DAG is a D-MAP, then, $I_{Pr} \Rightarrow I_G$ or $I_{Pr} \subseteq I_G$.
   For instance, following DAG is a D-MAP:



   $I_G = \{X_1 \perp X_2, X_3\}$
2. If a DAG is an I-MAP, then $I_G \Rightarrow I_{Pr}$ or $I_G \subseteq I_{Pr}$.
   For instance, following DAG is an I-MAP:



   $I_G = \{\}$

Problem 9: Bayesian networks (AD Exercise 4.17)

Prove that for strictly positive distributions, if B_1 and B_2 are Markov blankets for some variable X, then $B_1 \cap B_2$ is also a Markov blanket for X. [**Hint:** Appeal to the intersection axiom.]

Proof:

According to axiom 4.13 (intersection axiom):

$I_{\mathrm{Pr}}(X, Z \cup W, Y)$ and $I_{Pr}(X, Z \cup Y, W)$ only if $I_{Pr}(X, Z, Y \cup W)$.

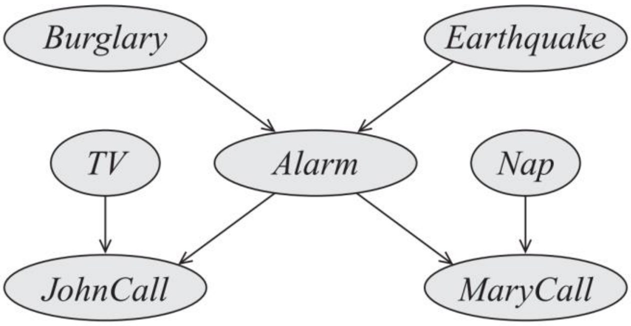
Let $Y = the\ whole\ set\ of\ variables$.

Then, $W_1 = Y - B_1 - X$, $W_2 = Y - B_2 - X$, $W_3 = Y - (B_1 \cap B_2) - X, B_0 = B_1 \cap B_2$.

Because $I_{Pr}(X, B_1, W_1)$ and $I_{\mathrm{Pr}}(X, B_2, W_2)$ holds, $I_{Pr}(X, B_0 \cup W_2, W_1)$ and $I_{Pr}(X, B_0 \cup W_1, W_2)$.

Thus, based on 4.13, $I_{Pr}(X, B_1 \cap B_2, Y - (B_1 \cap B_2) - X)$. Thus, $B_1 \cap B_2$ is also a Markov blanket for X.

Problem 10: Bayesian networks (Exercise 3.11 from KF)



Consider the Burglary Alarm network given above. Construct a Bayesian network over all the node **except** the Alarm that is a minimal I-map for the marginal distribution over the remaining variables (namely, over B, E, N, T, J, M). Be sure to get all the dependencies from the original network.

$dsep(B, \emptyset, \{E, T, N\})$
$dsep(E, \emptyset, \{B, T, N\})$
$dsep(T, \emptyset, \{B, E, N, M\})$
$dsep(N, \emptyset, \{B, E, T, J\})$
$dsep(J, \emptyset, N)$
$dsep(M, \emptyset, T)$
$dsep(B, E, \{T, N\})$
$dsep(B, J, N)$
$dsep(B, T, \{E, N\})$
$dsep(B, M, \{T\})$
$dsep(B, N, \{E, T\})$
$dsep(B, \{E, J\}, N)$
$dsep(B, \{E, T\}, N)$
$dsep(B, \{E, M\}, T)$
$dsep(B, \{J, T\}, N)$
$dsep(B, \{J, M\}, \emptyset)$

$dsep(B, \{J, N\}, \emptyset)$
$dsep(B, \{T, M\}, \emptyset)$
$dsep(B, \{T, N\}, \emptyset)$
$dsep(B, \{M, N\}, T)$
$dsep(B, \{T, M, N\}, \emptyset)$
$dsep(B, \{J, M, N\}, \emptyset)$
$dsep(B, \{J, T, N\}, \emptyset)$
$dsep(B, \{J, T, M\}, \emptyset)$
$dsep(B, \{E, M, N\}, T)$
$dsep(B, \{E, T, N\}, \emptyset)$
$dsep(B, \{E, T, M\}, \emptyset)$
$dsep(B, \{E, J, N\}, \emptyset)$
$dsep(B, \{E, J, M\}, \emptyset)$
$dsep(B, \{E, J, T\}, N)$
$dsep(B, \{J, T, M, N\}, \emptyset)$
$dsep(B, \{E, T, M, N\}, \emptyset)$
$dsep(B, \{E, J, M, N\}, \emptyset)$
$dsep(B, \{E, J, T, N\}, \emptyset)$
$dsep(B, \{E, J, T, M\}, \emptyset)$

$dsep(J, B, N)$

$dsep(J, E, N)$

$dsep(J, T, N)$

$dsep(J, N, \emptyset)$

$dsep(J, M, \emptyset)$

$dsep(J, \{B, E\}, N)$

$dsep(J, \{B, T\}, N)$

$dsep(J, \{E, T\}, N)$

$dsep(J, \{B, E, T\}, N)$

$dsep(T, B, \{E, N, M\})$

$dsep(T, E, \{B, N, M\})$

$dsep(T, J, \{M, N\})$

$dsep(T, M, \{B, E, N\})$

$dsep(T, N, \{B, E, M\})$

$dsep(T, \{B, E\}, \{M, N\})$

$dsep(T, \{B, J\}, N)$

$dsep(T, \{B, M\}, \{E, N\})$

$dsep(T, \{B, N\}, \{E, M\})$

$dsep(T, \{E, J\}, N)$

$dsep(T, \{E, M\}, \{B, N\})$

$dsep(T, \{E, N\}, \{B, M\})$

$dsep(T, \{J, M\}, N)$

$dsep(T, \{J, N\}, \emptyset)$

$dsep(T, \{M, N\}, \{B, E\})$

$dsep(T, \{J, M, N\}, \emptyset)$

$dsep(T, \{E, M, N\}, B)$

$dsep(T, \{E, J, N\}, \emptyset)$

$dsep(T, \{E, J, M\}, N)$

$dsep(T, \{B, M, N\}, E)$

$dsep(T, \{B, J, N\}, \emptyset)$

$dsep(T, \{B, J, M\}, N)$

$dsep(T, \{B, E, N\}, M)$

$dsep(T, \{B, E, M\}, N)$

$dsep(T, \{B, E, J\}, N)$

$dsep(T, \{E, J, M, N\}, \emptyset)$

$dsep(T, \{B, J, M, N\}, \emptyset)$

$dsep(T, \{B, E, M, N\}, \emptyset)$

$dsep(T, \{B, E, J, N\}, \emptyset)$

$dsep(T, \{B, E, J, M\}, N)$