Problem 1 (10 points.) Prove the chain rule. That is, for any probabilistic model composed of random variables $X_1, \ldots, X_n$ and any values $x_1, \ldots, x_n$, we have:

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i | x_1, \ldots, x_{i-1})$$

Proof by induction:

Base case: when n=1, $p(x_1)=p(x_1)$ which is obvious

Induction step: Let n=k-1, $p(x_1,\ldots,x_{k-1})=p(x_1)*p(x_2|x_1)*\ldots*p(x_{k-1}|\prod_{i=1}^{k-2} x_i) = \prod_{i=1}^{k-1} p(x_i|x_1,\ldots,x_{i-2})$

let n=k, $p(x_1,\ldots,x_k) = p((\prod_{i=1}^{k-1} x_i) \cap x_k)$

By applying the definition of conditional probability: $P[A\cap B]=P[A]P[B|A]$

$p(x_1,\ldots,x_k)= p((\prod_{i=1}^{k-1} x_i))*p(x_k|(\prod_{i=1}^{k-1} x_i)) = p(x_1)*p(x_2|x_1)*\ldots*p(x_{k-1}|\prod_{i=1}^{k-2} x_i)*p(x_k|(\prod_{i=1}^{k-1} x_i))$

The hypothesis holds for base case n=1, n=k-1, and n=k.

By induction, the chain rule has been proven.

Problem 2 (10 points.) Prove that the two definitions of conditional independence of random variables are equivalent. Let X, Y, Z be random variables. The two definitions are:

> **Definition 1:** X and Y are conditionally independent given Z if for any value x of X, any value y of Y, and any value z of Z, the following holds: $p(x, y|z) = p(x|z) \times p(y|z)$.
>
> **Definition 2:** X and Y are conditionally independent given Z if for any value x of X, any value y of Y, and any value z of Z, the following holds: $p(x|y, z) = p(x|z)$.

Definition 1: p(x,y|z) = p(x|z) * p(y|z)
Definition 2: p(x|y,z) = p(x|z)

def1: p(x,y|z) = $\frac{p(x,y,z)}{p(z)} = \frac{p(x|y,z)*p(y,z)}{p(z)} = p(x|y,z) * \frac{p(y,z)}{p(z)} = p(x|y,z) * p(y|z)$
We also know from def 1 that: $p(x|y,z) = p(x|z) * p(y|z)$
We get that $p(x|y,z) * p(y|z) = p(x|z) * p(y|z)$
So by dividing $p(y|z)$ from both sides, $p(x|y,z) = p(x|z)$ which is def 2
The equivalency of these two definitions has been shown.

Problem 3 (bonus question 10 points.) Let X, Y, Z be random variables. Prove or disprove the following statements. (That means, you need to either write down a formal proof, or give a counterexample.)

Statement 1 (5pts). If X and Y are (unconditionally) independent, is it true that X and Y are conditionally independent given Z?

Statement 2 (5pts). If X and Y are conditionally independent given Z, is it true that X and Y are (unconditionally) independent?

Statement 1 is FALSE, counterexample:

Let X and Y are two random variables that are independent. Let Z be a third random variable that Z=X+Y. Now, X and Y are not conditionally independent given Z.

Statement 2 is FALSE, counterexample:

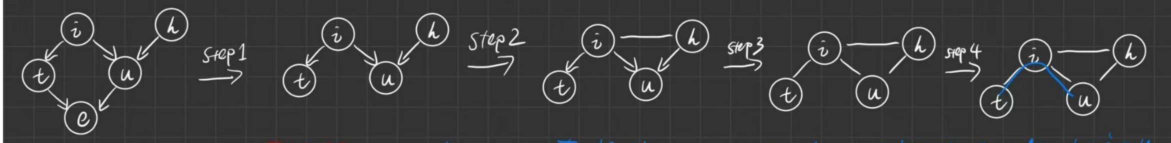
Assume we have 2 coins. A coin with a head and tail (HT), a coin with 2 head (HH).
I choose a random coin and flip it twice:
Two event defined here:
A= first coin toss results in an H
B= second coin toss results in an H.
C= Coin HT has been selected.

If given C, probabilities of event A and B occurs are conditionally independent.

However, event A and B are not unconditionally independent. If event A has occurred, the probability of we have chosen HH coin will be greater than we have chosen HT coin. As the result, increase the probability that B occurs. So, event A and B are dependent.

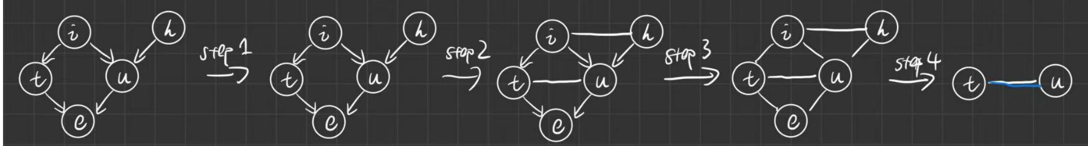
Problem 4 (40 points.) For the above Bayesian network, label the following statements about conditional independence as true or false. For this question, you should consider only the structure of the Bayesian network, not the specific probabilities. **Please show all four steps and the reasoning after the four steps taught in the class, otherwise no partial score will be given.**

1. $T \perp U$



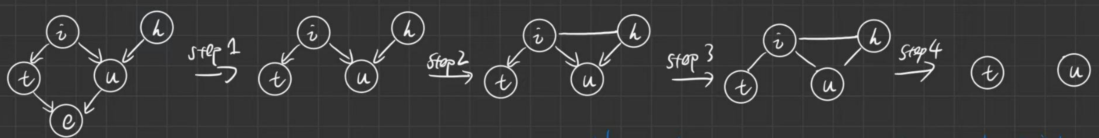
FALSE Conclusion: $T \not\perp U$ since there's a path between 2 nodes: $t - i - u$

2. $T \perp U \mid I, E, H$



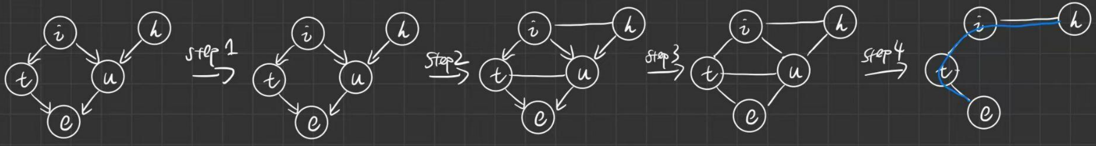
FALSE Conclusion: $T \not\perp U \mid I, E, H$ since there's a path between 2 nodes: $t - u$

3. $T \perp U \mid I, H$



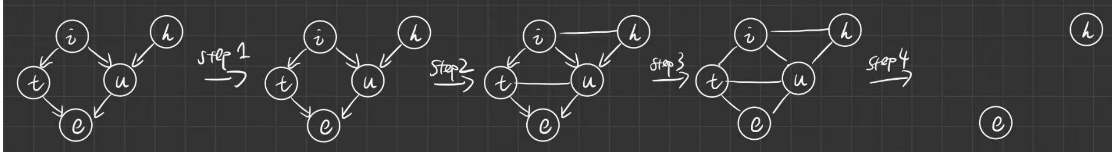
TRUE Conclusion: $T \perp U \mid I, H$ since no active path exist

4. $E \perp H \mid U$



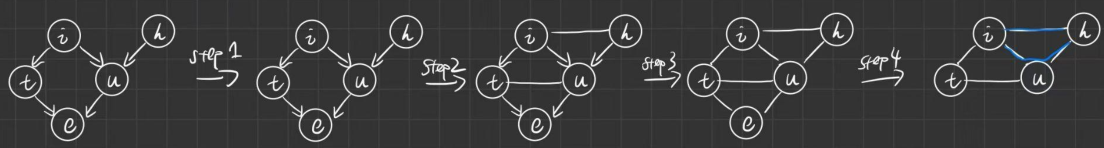
FALSE Conclusion: $E \not\perp H \mid U$ since $e - t - i - h$

5、 $E \perp H \mid U, I, T$



TRUE Conclusion: $E \perp H \mid U, I, T$ since no active path exist

6、 $I \perp H \mid E$

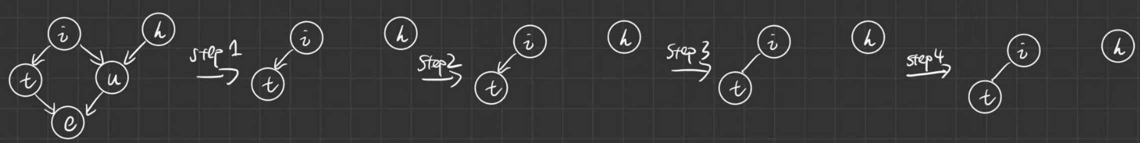


FALSE Conclusion: $I \not\perp H \mid E$ since $i - h$ or $i - u - h$

7、 $I \perp H \mid T$

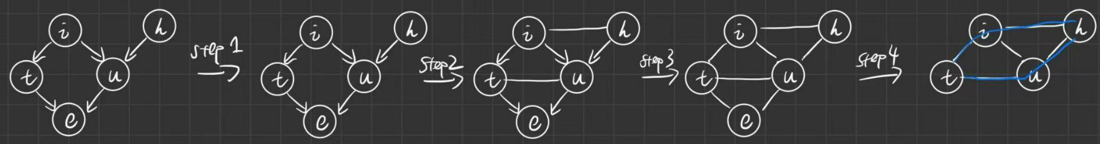
TRUE Conclusion: $I \perp H \mid T$ since no active path exist

8、 $T \perp H$



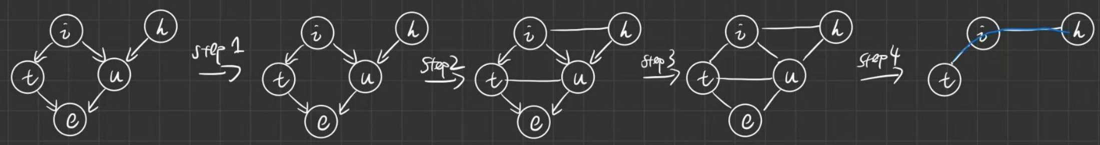
TRUE Conclusion: $T \perp H$ since no active path exist

9. $T \perp H \mid E$



FALSE Conclusion: $T \not\perp H \mid E$ since $t-i-h$ or $t-u-h$.

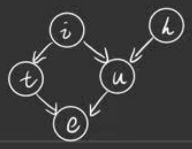
10. $T \perp H \mid E, U$



FALSE Conclusion: $T \not\perp H \mid E, U$ since $t-i-h$

Problem 5 (30 points).

Using variable elimination (by hand!), compute the probability that a student who did well on the test actually understood the material, that is, compute $P(+u|+e)$.



$$P(+u|+e) = \frac{P(+u,+e)}{P(+e)}$$

$$P(+u,+e) = \sum_{i,t,h} P(+u,+e,i,t,h)$$

$$= \sum_{i,t,h} P(i) \times P(t|i) \times P(+u|i,h) \times P(h) \times P(+e|t,u)$$

$$= \sum_{h} P(h) \sum_{i} P(i) \cdot P(+u|i,h) \sum_{t} P(t|i) \cdot P(+e|t,+u)$$

$f_2(h,i) = \sum_t P(t|i) \cdot P(+e|t,+u)$ eliminate t

$f_1(h) = \sum_i P(i) \cdot P(+u|i,h) \times f_2(h,i)$ eliminate i

$P(+u,+e) = \sum_h P(h) \times f_1(h)$

$[f_2(+h,+i) = f_2(-h,+i) = \sum_t P(t|+i) \cdot P(+e|t,+u)]$

$$f_2(\pm h,+i) = P(+t|+i) \cdot (+e|+t,+u) + P(-t|+i) \cdot P(+e|-t,+u)$$

$$= 0.8 \times 0.9 + 0.2 \times 0.7$$

$$= 0.86$$

$$f_2(\pm h,-i) = P(+t|-i) \cdot (+e|+t,+u) + P(-t|-i) \cdot P(+e|-t,+u)$$

$$= 0.5 \times 0.9 + 0.5 \times 0.7$$

$$= 0.8$$

$$f_1(+h) = \sum_i P(i) \cdot P(+u|i,+h) \cdot f_2(+h,i)$$

$$= P(+i) \cdot P(+u|+i,+h) \cdot f_2(+h,+i) + P(-i) \cdot P(+u|-i,+h) \cdot f_2(+h,-i)$$

$$= 0.7 \times 0.9 \times 0.86 + 0.3 \times 0.5 \times 0.8$$

$$= 0.6618$$

$$f_1(-h) = \sum_i P(i) \cdot P(+u|i,-h) \cdot f_2(-h,i)$$

$$= P(+i) \cdot P(+u|+i,-h) \cdot f_2(-h,+i) + P(-i) \cdot P(+u|-i,-h) \cdot f_2(-h,-i)$$

$$= 0.7 \times 0.3 \times 0.86 + 0.3 \times 0.1 \times 0.8$$

$$= 0.2046$$

$$P(+u,+e) = \sum_h P(h) \times f_1(h) = P(+h) \times f_1(+h) + P(-h) \times f_1(-h)$$

$$= 0.6 \times 0.6618 + 0.4 \times 0.2046$$

$$= 0.47892$$

$$\text{To calculate the denominator } P(+e) = \underbrace{P(+u,+e)}_{\text{Known}} + \underbrace{P(-u,+e)}_{\text{Unknown}}$$

we need to calculate $P(-u,+e)$

$$P(-u,+e) = \sum_{i,t,h} (-u,+e,i,t,h)$$
$$= \sum_h P(h) \sum_i P(i) \cdot P(-u|i,h) \sum_t P(t|i) \cdot P(+e|t,-u)$$

$$f_2(\pm h,+i) = P(+t|+i) \cdot (+e|+t,-u) + P(-t|+i) \cdot P(+e|-t,-u)$$
$$= 0.8 \times 0.5 + 0.2 \times 0.3$$
$$= 0.46$$

$$f_2(\pm h,-i) = P(+t|-i) \cdot (+e|+t,-u) + P(-t|-i) \cdot P(+e|-t,-u)$$
$$= 0.5 \times 0.5 + 0.5 \times 0.3$$
$$= 0.4$$

$$f_1(+h) = \sum_i P(i) \cdot P(-u|i,+h) \cdot f_2(+h,i)$$
$$= P(+i) \cdot P(-u|+i,+h) \cdot f_2(+h,+i) + P(-i) \cdot P(-u|-i,+h) \cdot f_2(+h,-i)$$
$$= 0.7 \times 0.1 \times 0.46 + 0.3 \times 0.5 \times 0.4$$
$$= 0.0922$$

$$f_1(-h) = \sum_i P(i) \cdot P(-u|i,-h) \cdot f_2(-h,i)$$
$$= P(+i) \cdot P(-u|+i,-h) \cdot f_2(-h,+i) + P(-i) \cdot P(-u|-i,-h) \cdot f_2(-h,-i)$$
$$= 0.7 \times 0.7 \times 0.46 + 0.3 \times 0.9 \times 0.4$$
$$= 0.3334$$

$$P(-u,+e) = \sum_h P(h) \times f_1(h) = P(+h) \times f_1(+h) + P(-h) \times f_1(-h)$$
$$= 0.6 \times 0.0922 + 0.4 \times 0.3334$$
$$= 0.18868$$

$$P(+e) = \underbrace{P(+u,+e)}_{\text{Known}} + \underbrace{P(-u,+e)}_{\text{Known}} = 0.47892 + 0.18868 = 0.6676$$

$$\boxed{P(+u|+e) = \frac{P(+u,+e)}{P(+e)} = \frac{0.47892}{0.6676} \approx 0.7174}$$