Yuchen (Bobby) Zhang, zhan2175
Prof. Bareinboim
CS 47100, Spring 2018
Purdue University
Apr 27, 2018

Homework 4

Problem 1

(a)

Let the whole cancer cases be the universe. Let E = diagnosed early, R = routine consultations. Given $P(E) = 0.7$, $P(R|E) = 0.6$, $P(\bar{R}|\bar{E}) = 0.9$, find $P(E|R)$.

$$
\begin{aligned}
P(E|R) &= \frac{P(E,R)}{P(R)} \\
&= \frac{P(R|E)P(E)}{P(R|E)P(E) + P(R|\bar{E})P(\bar{E})} \\
&= \frac{0.6 * 0.7}{0.6 * 0.7 + 0.1 * 0.3} \\
&= \frac{14}{15}
\end{aligned}
$$

(b)

$$
\begin{aligned}
\sum_{Z} P(x,z|y) &= \sum_{Z} \frac{P(x,y,z)}{P(y)} \\
&= \frac{\sum_{Z} P(x,y,z)}{P(y)} \\
&= \frac{P(x,y)}{P(y)} \\
&= P(x|y)
\end{aligned}
$$

Problem 2

(a)

There are 7 entries.

(b)

$P(a, b, c, d, e, f, g) = P(a)P(c|a, d)P(f|c)P(b)P(d|b)P(e|b)P(g|d, e)$

(c)

(i)

$P(g|A = 0) = P(A = 0)P(C|A = 0, D)P(F|C)P(B)P(D|B)P(E|B)P(g|D, E)$
$= \sum_E \sum_F \sum_C \sum_D \sum_B P(A = 0)P(C|A = 0, D)P(F|C)P(B)P(D|B)P(E|B)P(g|D, E)$
$= P(A = 0) \sum_E \sum_F \sum_C P(F|C) \sum_D P(C|A = 0, D)P(g|D, E) \sum_B P(B)P(D|B)P(E|B)$

(ii)

Assume all variables are binary.
$P(g|A = 0) = P(A = 0) \sum_E \sum_F \sum_C P(F|C) \sum_D P(C|A = 0, D)P(g|D, E) \sum_B P(B)P(D|B)P(E|B)$
$= P(A = 0) \sum_E \sum_F \sum_C f_1(C, F) \sum_D f_2(C, D)f_3(D, E) \sum_B f_4(B)f_5(B, D)f_6(B, E)$
$= P(A = 0) \sum_E \sum_F \sum_C f_1(C, F) \sum_D f_2(C, D)f_3(D, E)f_7(D, E)$
$= P(A = 0) \sum_E \sum_F \sum_C f_1(C, F)f_8(C, E)$
$= P(A = 0) \sum_E \sum_F f_9(E, F)$
$= P(A = 0) \sum_E f_{10}(E)$
During each steps of the computation:
1. $f_7(D, E)$ has factor of $2^2 = 4$
2. $f_8(C, E)$ has factor of $2^2 = 4$
3. $f_9(E, F)$ has factor of $2^2 = 4$

So the size of the largest factor is 4.

We could try eliminate F first by $\sum_F f_1(C, F) = f_7(C)$. Then we eliminate C by $\sum_C f_2(C, D)f_7(C) = f_8(D)$. Then eliminate D by $\sum_D f_3(D, E)f_5(B, D)f_8(D) = f_9(B, E) = f_6(B, E)$, so no need to re-calculate the value. Then eliminate E by $\sum_E f_6(B, E) = f_{10}(B) = f_4(B)$. In this case, the size of the factor will be $2^1 = 2$. The ordering is thus F, C, D, E, B.

(d)

(i)

$P(f|b) = P(A)P(C|A,D)P(f|C)P(b)P(D|b)P(E|b)P(G|D,E)$
$= \sum_G \sum_E \sum_D \sum_C \sum_A P(A)P(C|A,D)P(f|C)P(b)P(D|b)P(E|b)P(G|D,E)$
$= P(b) \sum_G \sum_E P(E|b) \sum_D P(D|b)P(G|D,E) \sum_C P(f|C) \sum_A P(A)P(C|A,D)$

(ii)

Assume all variables are binary.

$P(f|b) = P(b) \sum_G \sum_E P(E|b) \sum_D P(D|b)P(G|D,E) \sum_C P(f|C) \sum_A P(A)P(C|A,D)$
$= P(b) \sum_G \sum_E f_1(E) \sum_D f_2(D) f_3(D,E,G) \sum_C f_4(C) \sum_A f_5(A) f_6(A,C,D)$
$= P(b) \sum_G \sum_E f_1(E) \sum_D f_2(D) f_3(D,E,G) \sum_C f_4(C) f_7(C,D)$
$= P(b) \sum_G \sum_E f_1(E) \sum_D f_2(D) f_3(D,E,G) f_8(D)$
$= P(b) \sum_G \sum_E f_1(E) f_9(E,G)$
$= P(b) \sum_G f_{10}(G)$

During each steps of the computation:
1. $f_7(C,D)$ has factor of $2^2 = 4$
2. $f_8(D)$ has factor of $2^1 = 2$
3. $f_9(E,G)$ has factor of $2^2 = 4$

So the size of the largest factor is 4.

There is no ordering such that the size of the largest factor is smaller because $f_3(D,E,G)$ and $f_6(A,C,D)$ has three variables and cannot be summed out two variables at once.

Problem 3

(a)

Since no independence information is given, we can only transform the original expression to $P(y|x_1, x_2) = \frac{P(x_1,x_2,y)}{P(x_1,x_2)} = \frac{P(x_1,x_2|y)P(y)}{P(x_1,x_2)}$. Only the second set is sufficient.

(b)

Since $(X_1 \perp\!\!\!\perp X_2|Y)$, we can further expand the expression to be $\frac{P(x_1|y)P(x_2|y)P(y)}{P(x_1,x_2)}$. In this case, both the first and the second sets are sufficient.

Problem 4

(a)

The variables C, E, F, H, I are d-connected to A given S = B.

By causal chain, we could easily know that H, F, C are d-connected to A. By common cause, we know I is d-connected to A. By causal chain again, we know E is also connected to A.

Since the G and H has a common effect I, which blocks the connectivity of G and H, G and D are thus separated from A. B is given, thus separated from A.

(b)

The variables B, D, E, G, H, I are d-connected to A given S = J.

Since J is the descendent of A and D, D and G are connected to A. By casual chain, B is connected to A. By common cause, E, I, H are connected to A. H is connected to A since both (H,D,G,A) and (H,E,B,D,G,A) works.

Since E and F has a common effect I, which blocks the connectivity of E and F, C and F are thus separated from A. J is given, thus separated from A.