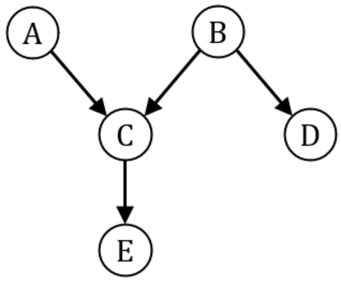
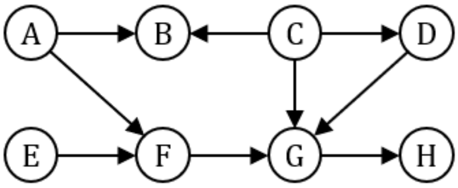
PGM-Assignment #2

1. Consider the following Bayesian Network consisting of four **binary** variables A, B, C, D, E:



Please derive the formula for calculating $P(D=1, E=0|A=1)$ according to the structure of BN. (**Note**: each term in your final expression should be in the form of $P(child|parents)$, such as $P(A), P(C|A,B)$.)

2. For the following Bayesian Network:



1) If the above graph is an I-Map of a distribution P, please write down the factorized representation of P according to the graph.

2) Please determine whether the following independences are true.

a) $D \perp F$

b) $A \perp G \mid F$

c) $A \perp D \mid B$

d) $D \perp E \mid H$

e) $A \perp C \mid H$

f) $A \perp H \mid B, F$

3) What is the largest set S_1, S_2, S_3 for which following independence relationships hold?

a) $A \perp S_1$

b) $E \perp S_2 \mid G$

c) $D \perp S_3 \mid C$

3. Now we have four binary variables A, B, C, D. Following are two cases of independent and conditionally independent relations. According to them, please draw a minimal I-map for each case and check whether the I-map is perfect. If not, could you find a perfect map for each of them?

(1) $(A \perp C)$, $(B \perp D \mid A, C)$

(2) $(B \perp C \mid A, D)$, $(A \perp D \mid B, C)$

4. [Modified from Joris Mooij, 2018] Suppose we collect electronic patient records to investigate the effectiveness of a new drug against a certain disease. Patients were divided into two groups: those that took the drug (treatment group), and those that didn't take the drug (control group). Some of the patients recovered, others unfortunately didn't.

	Recovery	No recovery	Total
No drug	47	13	60
Drug	49	11	60
Total	96	24	120

(1) You find the following experimental data.

Please compare p(recovery |drug) with p(recovery|no drug) and answer "do you think the drug is effective?"

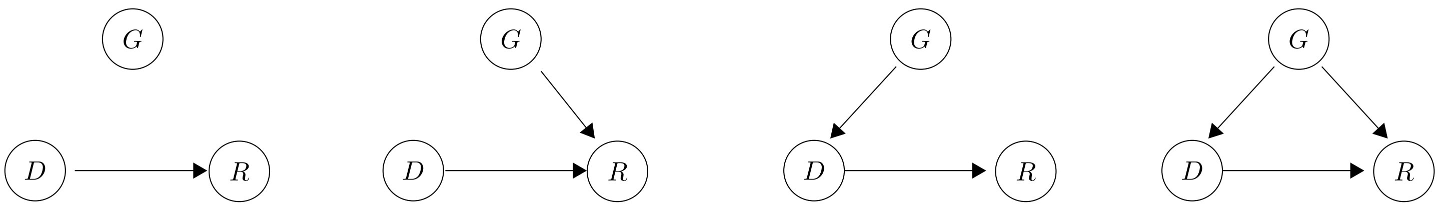
(2) Upon closer inspection of the records, you noticed you could further group patient by gender:

Females	Recovery	No recovery	Total
No drug	19	1	20
Drug	37	3	40
Total	56	4	60

Males	Recovery	No recovery	Total
No drug	28	12	40
Drug	12	8	20
Total	40	20	60

Please compare both p(recovery | drug, male) with p(recovery |no drug, male) and p(recovery | drug, female) with p(recovery | no drug, female). Answer again "do you think the drug is effective?"

(3) Next, we use random variable G to indicate the gender $(0 : \text{female}, 1 : \text{male})$, D to indicate whether using the drug $(0 : \text{No drug}, 1 : \text{Drug})$, and R to indicate recovery or not $(0 : \text{no recovery}, 1 : \text{recovery})$. Based on the above **observations**, justify the Bayesian network model (d) from the following is more reasonable.



(a) (b) (c) (d)

There is one important detail here. In the data generating process, we **set** the variable D as 0 or 1. But we model this as we **observe** it. To formalize this, we use $p(Y = y|X = x)$ to represent what the distribution of Y is given that I **observe** variable X takes value x, and $p(Y = y|\mathbf{do}(x))$ to represent what the distribution of Y is if I were to **set** the value of X to x. The former is an **observational** problem while the latter is **interventional**.

Return to our problem. Now you should confirm yourself what we really want is $p(\text{recovery}|\mathbf{do}(\text{drug}), \cdots)$ or $p(R = 1|\mathbf{do}(D = 1), \cdots)$. Assume we use the (d) model above. (Usually, G is called as a confounder variable.) We further make an important assumption: all edges in the graph indicate <u>**direct causal effects**</u>. Based on these, **please calculate :**

$$p(R = 1|\mathbf{do}(D = 1))$$
$$p(R = 1|\mathbf{do}(D = 0))$$
$$p(R = 1|\mathbf{do}(D = 1), G = 0)$$
$$p(R = 1|\mathbf{do}(D = 0), G = 0)$$
$$p(R = 1|\mathbf{do}(D = 1), G = 1)$$
$$p(R = 1|\mathbf{do}(D = 0), G = 1)$$

And answer again: "do you think the drug is effective?".

(Hint:

- As a toy example, consider Y: the car speed following $p(Y)$. Let X denote the speed by a detector which follows a conditional distribution $p(X|Y)$

  There are two cases: we (a) **observe** $x = 60$, (b) **set** $x = 60$. You can figure out the distribution of Y for case (a) in form of $p(Y|X = 60)$. For case (b), setting X artifically has no effect on Y, so the distribution of Y is still $p(Y)$.

- A **do** operation temporarily removes the direct link between G and D generating a mutilated graph.)

5. Human computer interaction (HCI) studies the design and use of computer technology, and focuses on the interfaces between people (users) and computers. The ubiquitous graphical interface, the mouse and windows, which are familiar to us, are all cases of HCI. Besides the traditional HCI, much attention has been paid to human-centered interaction technologies. Speaking and posture, as the main ways we human express ourselves, are taken as new interaction methods, which will help us to easily order and control the computers.

Here is a HCI system which can recognize human's instructions based on his gesture (by a camera) and voice (by a microphone). The raw signals detected by the two sensors should be treated as the observations of the real gestures and voices. The computer samples the sensors 10 times per second. According to these descriptions:

(1) Please draw a Bayesian network to model the human instruction recognition process at a given time.

(2) Please extend your model to consider the continuous sampling process across a period of time.

(You can use the symbols you like to represent the essential variables you think, but please explain the meaning each variable symbol stands for. We recommend your model can show the generative process.)

6. (Optional) In the question 4, we try to understand the meaning of **"intervention"**, which is the second level of causal inference. Here, we use a simple example to consider the third level, **"counterfactuals"**, which means mining worlds that could have been.

[Modified from Judea Pearl, 2018] Suppose that we are working to analyze the effects of education and work experience on employee's salary. We have collected some data as the following table shown. For each employee u, we use the work years to measure his/her experience $Ex(u)$, and use the level $0, 1, 2$ to denote the education $Ed(u)$, where $0 = \textit{bachelor degree}$, $1 = \textit{master degree}$, and $2 = \textit{doctor degree}$. For the salary S, we use $S_0(u)$ to represent the salary of employee u if u have bachelor degree but not master degree ($S_1(u)$ and $S_2(u)$ have similar meanings).

Employee u	$Ex(u)/year$	$Ed(u)$	$S_0(u)$	$S_1(u)$	$S_2(u)$
Alice	3	0	8,000	?	?
Bob	2	1	?	12,000	?
Carla	2	2	?	?	15,000
David	2	1	?	11,000	?
Ellen	0	2	?	?	13,000
Frank	4	0	10,000	?	?
...					

Then we consider some examples of imputing the missing data (denoted by "?" in the table):

a) If Bob had doctor degree, how much salary would he get?

b) If Alice had master degree, how much salary would she get?

c) If Ellen hadn't study for the doctor degree and start working after getting master degree (she would have 3 years work experience), how much salary would she get?

These examples are all about hypotheses, which means we couldn't change the things had happened and could only image them to make inference.

So, please consider the following questions and give your results.

(1) What causal relationships between Ex, Ed, and salary S do you think? Could you please use a Bayesian network to represent them? (Just explain it based on your understanding for this application scenario, and don't need to calculate the independence relationship.)

(2) (Open question) Do you have some ideas on how to answer the counterfactuals inference like the examples above?