Homework #1

Due: 17.03.2020

Problem 1. In the kitchen in your apartment, you put all your 10 forks in the left drawer and all 10 knives in the right drawer. Your roommate, who does not agree with your organizational approach, comes in, takes two forks from the left drawer and tosses them into the right drawer. She then takes at random an item (knife or fork) from the right drawer and tosses it in the left drawer. After this exchange, you come in and randomly pick up an item from a randomly chosen drawer. Given you have picked up a knife, what is the probability that you have opened the left drawer?

Problem 2. Consider a Gaussian random vector $X \sim N(\mu, \Sigma)$, where $\mu = [1\ 5\ 2]^T$ and

$$\Sigma = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

Find the pdfs of

(a) X_1

(b) $2X_1 + X_2 + X_3$

(c) X_3 given(X_1, X_2)

(d) (X_2, X_3) given X_1

Problem 3. A player throws a fair die and simultaneously flips a fair coin. If the coin lands heads, then she wins twice, and if tails, the one-half of the value that appears on the die. Explain her expected winnings.

Problem 4. Prove or disprove (by providing a counter-example) :

$$P(X, Y|Z) = P(X|Y, Z)P(Y|Z).$$

Problem 5. Suppose we wish to calculate $P(H|E_1, E_2)$, and we have no conditional independence information. Which of the following sets of numbers are sufficient for the calculation?

(a) $P(E_1, E_2), P(H), P(E_1|H), P(E_2|H)$

(b) $P(E_1, E_2), P(H), P(E_1, E_2|H)$

(c) $P(E_1|H), P(E_2|H), P(H)$

.

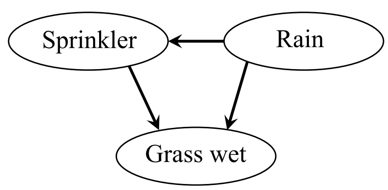

Figure 1:

Problem 6. Suppose the grass could be wet either because the sprinkler is on or it's raining. Moreover, assume that rain directly affects the use of sprinkler (You won't turn on the sprinkler if it is raining). Figure 1 shows the corresponding Bayesian network. Let the following tables capture the joint probability mass functions of the events as well as some marginals.

Rain T	Rain F
0.2	0.8

Rain	Sprinkler T	Sprinkler F
F	0.4	0.6
T	0.01	0.99

Sprinkler	Rain	Grass Wet T	Grass Wet F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

(a) What is the probability that it is raining, given that the grass is wet?

(b) Now consider the following interventional question. What is the probability that it would rain, given that we wet the grass?

(c) Is set $Z = \{Rain\}$ a valid adjustment set for predicting the effect of sprinkler is turned on (the do operation in 2nd part) on grass being wet?

Problem 7. For two discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, mutual information is defined as

$$I(X;Y) = \sum_{(x,y)\in(\mathcal{X},\mathcal{Y})} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)},$$

where $P_{X,Y}$ is the joint probability distribution of X and Y, and P_X and P_Y are the marginal distributions. $\mathcal{X}$ and $\mathcal{Y}$ are the sets the random variables take their values in (called the alphabet). A large mutual information between two random variables implies that one could be used as a good predictor for the other.
Now consider the following SEM:

$$S: \quad \begin{aligned} A &= N_A \\ H &= A \oplus N_H \\ B &= H \oplus N_B \end{aligned}$$

with graph

$$A \to H \to B$$

where $N_A \sim Ber(1/2)$, $N_H \sim Ber(1/3)$ and $N_B \sim Ber(1/20)$ are independent. The symbol $\oplus$ denotes addition modulo 2 (i.e. $1 \oplus 1 = 0$).
(a) Between A and B, which one is a better predictor for H? (use mutual information as a measure)
(b) Now lets say we where interested in learning whether A or B are causing H. Use use interventional distributions to figure out the causal structure.
(c) Can you conclude from parts (a) and (b) that a larger influence implies causation or the statement is wrong?