Mathematical Underpinnings Lab 6 03.04.2024

Task 1

Let $Z = (Z_1, Z_2, \dots, Z_s)$ and $S = \{1, 2, \dots, s\}$. We define feature selection criteria

$$SECMI2(X,Y|Z) = CIFE(X,Y|Z) = I(X;Y) + \sum_{i \in S} II(Y;X;Z_i),$$

and

$$SECMI3 = CIFE3(X,Y|Z) = I(X;Y) + \sum_{i \in S} II(Y;X;Z_i) + \sum_{i<j, i,j \in S} II(Y;X;Z_i;Z_j).$$

a) Write a function performing conditional independence test based on conditional permutations.
Input: X, Y, Z (Z might be multivariate), B (default B=50), stat (one of 'CMI', 'SECMI2' and 'SECMI3')
Output: test statistic, p-value

b) Sample $n = 100$ observations from the model shown in Figure 1. Use the same distributions and discretization schemes as in Lab 5 Task 2 (first normal distribution, then discretization into two bins: > 0 and < 0). Run tests for hypotheses:

$$H_0^{(1)} : X \perp\!\!\!\perp Y | (Z_1, Z_2) \quad \text{and} \quad H_0^{(2)} : X \perp\!\!\!\perp Y | (Z_2, Z_3).$$

Which of these conditional independencies is true?
Run both the experiments $N = 100$ times. How many times the null hypotheses were rejected in each case?

c) Sample $n = 100$ observations from the following distribution:
The distribution of Y is defined as follows:

$$P(Y = 1 | X + Z_1 + Z_2 =_2 1) = P(Y = 0 | X + Z_1 + Z_2 =_2 0) = 0.8,$$

and

$$P(Y = 0 | X + Z_1 + Z_2 =_2 1) = P(Y = 1 | X + Z_1 + Z_2 =_2 0) = 0.2,$$

where $=_2$ denotes addition modulo 2. There is also a variable Z_3 independent of (X, Y, Z_1, Z_2) . All variables X, Z_1, Z_2, Z_3 are independent and binary with the probability of success equal to 0.5.
Run tests for hypotheses:

$$H_0^{(1)} : X \perp\!\!\!\perp Y | (Z_1, Z_2) \quad \text{and} \quad H_0^{(2)} : X \perp\!\!\!\perp Y | (Z_2, Z_3).$$

In which case the variables are conditional independent? Run both the experiments $N = 100$ times. How many times the null hypotheses were rejected in each case?

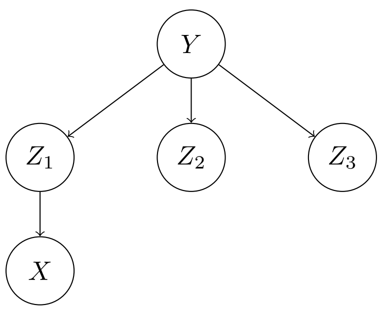

Figure 1: Model (X, Y, Z_1, Z_2, Z_3)

Task 2

In this task, use R and `bnlearn` package.

a) Load the bayesian network called `alarm` from Bayesian network repository.

b) Draw its graphical representation. Check Markov Blanket for `HR` variable.

c) Draw a sample from the `alarm` network ($n = 1000$).

d) Learn the Markov blanket of `HR`. Use varius methods and tests of conditioanl independence (`learn.mb`, methods: 'gs', 'iamb', tests: 'mi', 'mc-mi', 'sp-mi'). Compare the results with the true Markov blanket. Use TPR and FDR (definitions can be found here).

e) Check the performance of all the methods and tests for `HR` node.