Mathematical Underpinnings Lab 5 27.03.2024

Task 1 (Tests of conditional independence)

Asymptotic test of conditional independence

$$H_0 : X \perp\!\!\!\perp Y|Z.$$

based on conditional mutual information uses the fact that

$$\text{if } X \perp\!\!\!\perp Y|Z, \text{ then } 2n\widehat{CMI} \to \chi^2_{(|\mathcal{X}|-1)(|\mathcal{Y}|-1)|\mathcal{Z}|}.$$

Conditional permutation test consists in computing $C\hat{M}I_b$ based on a sample, in which X is permuted separately for each distinct value of Z and then the p-value is computed in the following way

$$\frac{1 + \sum_{b=1}^{B} \mathbb{I}(\widehat{CMI} \leq \widehat{CMI}_b^*)}{1 + B}.$$

a) Write a function which runs asymptotic conditional independence tests.
Input: X, Y, Z, stats (one of 'cmi' - based on conditional mutual information, 'chi2' - Pearson's statistic).
Output: test statistic value, p-value

b) Write a function for conditional permutations of X. They consists in permuting X for each distinct value of Z separately, so the distributions of (X, Z) and (Y, Z) remain unchanged while the conditional dependence between X and Y given Z vanishes.
Next, write a function performing conditional independence test based on permutations.
Input: X, Y, Z, B - number of permutations
Output: test statistic value, p-value

c) Draw a sample from a distribution, in which X and Y are conditionally dependent given Z and a sample, in which X and Y are conditionally independent. Next, test for independence using:

- asymptotic test of conditional independence based on $\hat{I}(X, Y|Z)$,
- conditional permutation test.

Task 1 (MI vs CMI)

In this task we will focus on the answering the questions: Are X and Y independent? Are X and Y conditionally independent given Z?

In the following, we will denote continuous variables using notation with tilde e.g. $\tilde{Z}$ and their discretized versions by e.g. Z.

We consider 3 models (see Figures 1a-1c). We sample data in the following way (number of observations $n = 1000$):

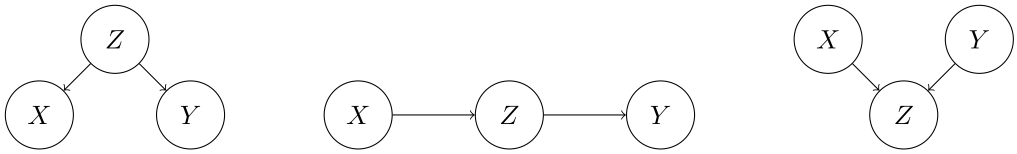
- **Model 1** Draw $\tilde{Z} \sim \mathcal{N}(0, 1)$, discretize $\tilde{Z}$ into two bins: $\tilde{Z} < 0$ ($Z = -1$) and $\tilde{Z} \geq 0$ ($Z = 1$). Draw independently $\tilde{X} \sim \mathcal{N}(Z/2, 1)$ and $\tilde{Y} \sim \mathcal{N}(Z/2, 1)$ and discretize $\tilde{X}$ and $\tilde{Y}$ in the same way as $\tilde{Z}$.
- **Model 2** Draw $\tilde{X} \sim \mathcal{N}(0, 1)$, discretize $\tilde{X}$, draw $\tilde{Z} \sim \mathcal{N}(X/2, 1)$, discretize $\tilde{Z}$, and in the same way draw $\tilde{Y}$ (i.e. $\tilde{Y} \sim N(Z/2, 1)$) and then discretize $\tilde{Y}$ to obtain Y.
- **Model 3** Draw independently $\tilde{X} \sim \mathcal{N}(0, 1)$ and $\tilde{Y} \sim \mathcal{N}(0, 1)$, discretize $\tilde{X}$ and $\tilde{Y}$ as previously, draw $\tilde{Z} \sim \mathcal{N}((X + Y)/2, 1)$, discretize $\tilde{Z}$.

a) In which models is there independence or conditional independence between X and Y?

b) For models 1-3 compute mutual information $I(X, Y)$ and conditional mutual information $I(X, Y|Z)$.

c) Run tests of independence and conditional independence.

d) * Question: why we first discretize the variable and then define the ensuing variables using the discretized version of the variable as e.q. in model 1 we first discretize $\tilde{Z}$ to obtain Z and then sample X and Y using Z? How will the answers to the questions about the dependence and conditional dependence change if we first sample all the variables and then discretize them all at the end? A hint: Figure 2.



(a) Model 1 (b) Model 2 (c) Model 3

Figure 1: Models

(a) Model 1 - discretization before sampling (b) Model 1 - discretization at the end

Figure 2: Discretization