Mathematical Underpinnings Lab 4 20.03.2024

Task 1 (Correlation vs MI)

a) Sample $n = 1000$ observations from multivariate normal distribution:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\rho \in [0, 1)$.

- Draw a heatmap of a distribution for $\rho = 0, 0.5, 1$.
- Compute Pearson correlation of X and Y, and mutual information (after discretizing variables into 10 bins each) for variuos values of ρ.
- For each ρ repeat the experiment from the previous bullet point $N = 100$ times. Visualise the results.
- Draw plots of $\hat{I}(X, Y)$ as a function of $-\log(1-\rho^2))/2$ and $-\log(1-\hat{\rho}^2))/2$ (use $\hat{I}(X, Y)$ and $\hat{\rho}$ from the previous bullet point). What is the relationship between x and y in the plots?

b) Sample $n = 1000$ observations from a normal distribution $X \sim \mathcal{N}(0, 1)$ and $Y = X^2 + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

- Draw a heatmap of a distribution for $\sigma^2 = 0, 0.5, 2$.
- Compute Pearson correlation of X and Y, and mutual information (after discretizing variables into 10 bins each) for various values of σ^2.
- For each σ^2 repeat the experiment $N = 100$ times. Visualise the results.

Task 2 (Tests of independence)

Asymptotic tests of independence might be based on mutual information or Pearson's χ^2 statistic. We want to test the following hypothesis

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

In this approach we compute $\widehat{MI} = \hat{I}(X, Y)$ or χ^2 for X and Y and then, using the fact that

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

and

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

we compute p-values using values of the cumulative distribution function of $\chi^2_{(\mathcal{X}-1)(\mathcal{Y}-1)}$.

In **permutation tests** we compute B times the value of the test statistics for a sample, in which X is permuted ($\widehat{MI}_b$ for $b = 1, 2, \ldots, B$) and then we compare $\widehat{MI}$ with $\widehat{MI}_b$ using the following formula for p-value:

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

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

b) Write a function performing independence test based on permutations.
Input: X, Y, B - number of permutations
Output: test statistic value, p-value
In a function, mutual information $\widehat{MI}_b^*$ should be computed for resampled samples (X_b^*, Y) for $B = 1, 2, \ldots, B$, where X_b^* is a random permutation of X.
Useful functions: `np.random.permutation`

c) Draw a sample from a distribution, in which X and Y are dependent and a sample, in which X and Y are independent (you may use the example from task 1 a)). Next, test for independence using:

- asymptotic test based on $2n\hat{I}(X, Y)$,
- Pearson's chi-squared test,
- permutation test.