1. Let p be the equiprobable distribution on natural numbers from 1 to 3, and q be the equiprobable distribution on natural numbers from 1 to 6.

   Calculate $CE(q||p)$ and $CE(p||q)$.

2. The random variable Y_i are independent identically distributed and take the values $1, 2$ and 3 with unknown probabilities $p_1 + p_2 + p_3 = 1$.

   You have a sample of 1000 observations with $\bar{y} = 2$. The law of large numbers allows you to assume that $\mathbb{E}(Y_i) = 2$.

   Estimate $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_3$ if you believe in the most unpredictable world given observed constraints.

   Hint: you should maximize the entropy of Y_i given that $\mathbb{E}(Y_i) = 2$.

3. We play a game. I flip a fair coin. If the coin shows head then you get your stake doubled with probability 0.5 and you get nothing with probability 0.5. If the coin shows tail then you get your stake doubled with probability 0.9 and you get nothing with probability 0.1.

   You maximize long-term interest rate.

   (a) How much of your current fortune your should invest if you make the stake *before* the coin toss?

   (b) How much of your current fortune your should invest if you make the stake *after* the coin toss and you know the result of the toss?

4. The response variable is binary. Elon Musk minimizes the Gini impurity index and splits the node of a tree into two non-empty child nodes.

   Provide an example for each case or prove that the case is not possible:

   (a) The impurity of both child nodes is lower than the impurity of the original node.

   (b) The impurity of the left child node is lower than the impurity of the original node, and the impurity of the right child node is higher.

5. Elon Musk forecasts the scalar random variable y using the ensemble of n models. The model number i provides its own forecast $\hat{y}_i$ with $\mathbb{C}\mathrm{orr}(y, \hat{y}_i) = 0.1$. Models are trained on random subsets of the initial dataset and hence their forecasts are correlated, $\mathbb{C}\mathrm{orr}(\hat{y}_i, \hat{y}_j) = 0.2$ for $i \neq j$. The variances of the forecasts of each model are equal, $\mathbb{V}\mathrm{ar}(\hat{y}_1) = \mathbb{V}\mathrm{ar}(\hat{y}_2) = \ldots$

   Elon Musk considers the ensemble forecast $\hat{y} = \sum \hat{y}_i / n$.

   (a) Find $\mathbb{C}\mathrm{orr}(\hat{y}, y)$.

   (b) Find the limit $\mathbb{C}\mathrm{orr}(\hat{y}, y)$ when $n \to \infty$.

   Hint: please be careful, in this problem i is the number of a model and not the number of an observation.

6. Consider the naive bootstrap procedure. The initial sample consists of n observations. Consider the random variable N_1 — the number of copies of the first initial observation in a bootstrap sample.

   Assume that the number of observations n is very very big.

   (a) Find $\mathbb{P}(N_1 = 0)$ and $\mathbb{P}(N_1 = 1)$.

   (b) Find the general formula $\mathbb{P}(N_1 = k)$.

   Hint: you know the name of this distribution :)