# Assignment 2

**Submission deadline:** last classes before **16.03.2017**

**Points: 13 + 1bonus**

The following assignment is intended to provide a solid foundation for the mathematics required to comprehend the course.

Please write the solutions to the theoretical problems on paper and hand them to the instructor.

Please submit the coding exercises in the form of an IPython notebook.

## 1 Puzzles

The following two problems are best solved by writing down all possible events and their probabilities, and then removing the events that are impossible under the constraints described in the text of the assignments.

**Problem 1.** **[1p]** Think hard and think twice

You're a guest star in a popular TV show *Go All the Way*. You have won the episode, and it's your turn to draw the prize. There are three gates you can choose from, yet only one of them holds the grand prize (the other two are empty and leave you with empty hands). The prize is assigned to one of the gates with probability $\frac{1}{3}$.

As 1 has always been your lucky number, you have chosen gate No. 1. The host shows you the content of gate No. 2, which you haven't selected. It is empty. He offers you to switch to gate No. 3. Will switching the gate increase the probability of winning the grand prize?

**Problem 2.** **[1p]** If you want the right answers, ask the right questions

A friend of yours has two children. Assuming that conceiving a boy is equally probable as conceiving a girl, we can assume a priori that he has both a boy and a girl (or a girl and a boy) with probability $\frac{1}{2}$, two boys with probability $\frac{1}{4}$ and two girls with $\frac{1}{4}$.

(a) What is the probability that he has a daughter, if you know that he has at least one son?

(b) What is the probability that he has a daughter, if just a while ago you have seen his son playing in the garden? Equivalently: you know his son is the younger child.

## 2 Bayes' theorem

**Problem 3.** **[2p]** Bayes' theorem

Bayes' theorem allows to reason about conditional probabilities of causes and their effects:

$$p(A, B) = p(A|B)p(B) = p(B|A)p(A) \tag{1}$$

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{2}$$

Bayes' theorem allows us to reason about probabilities of causes, when we observe their results. Instead of directly answering the hard question $p(\text{cause}|\text{result})$ we can instead separately work out the marginal probabilities of causes $p(\text{cause})$ and carefully study their effects $p(\text{effect}|\text{cause})$.

Solve the following using Bayes' theorem.

(a) **[1p]** There are two boxes on the table: box #1 holds two black balls and eight red ones, box #2 holds 5 black ones and 5 red ones. We pick a box at random (with equal probabilities), and then a ball from that box. What is the probability, that the ball came from box #1 if we happened to pick a red ball?

(b) **[1p]** The government has started a preventive program of mandatory tests for the Ebola virus. Mass testing method is imprecise, yielding 1% of false positives (healthy, but the test indicates the virus) and 1% of false negatives ( having the virus but healthy according to test results). As Ebola is rather infrequent, lets assume that it occurs in one in a million people in Europe. What is the probability, that a random European, who has been tested positive for Ebola virus, is indeed a carrier?

Suppose we have an additional information, that the person has just arrived from a country where one in a thousand people is a carrier. How much will be the increase in probability?

How accurate should be the test, for a 80% probability of true positive in a European?

**Problem 4.** **[3p + 1bp]** Naive Bayes classifier

The Bayes theorem allows us to construct a classifier in which we model how the data are generated. Here we will describe a very simple and popular example of such a classifier called the naïve Bayes classifier. Despite its simplicity It is quite effective for classification of text documents (e.g. as spam and non-spam).

We will model the generation of text documents (treated as sequences of words) $D = W_1, W_2, \ldots, W_n$ in a two stage process. First, document class $C_j$ is drawn at random with probability $p(C_j)$, also called the *a priori probability*.

Next we need to define the class-conditional probability $p(D|C_j)$. We will make the following simplifying (naïve) assumption: every word in the document is drawn independently at random with the probability $p(W_i|C)$. In other words, we assume that:

$$p(D|C_j) = p(W_1, W_2, \ldots, W_n|C_j) \approx p(W_1|C_j)p(W_2|C_j)\ldots p(W_n|C_j).$$

To infer the class of a document (classify it) we apply the Bayes theorem:

$$p(C_j|D) = \frac{p(D|C_j)p(C_j)}{p(D)} = \frac{p(C_j)p(W_1|C_j)p(W_2|C_j)\ldots p(W_n|C_j)}{p(D)}.$$

Please note that since we assumed only a finite number of classes, we can compute the term $p(D)$ by making sure that the *a posteriori probabilities* $p(C_j|D)$ sum to 1 over all classes.

In this exercise we will try to mimic the language-guessing feature of Google Translate (`https://translate.google.com/`), although on a much smaller scale. We are given an input which is a lower-case sequence of characters (such as **some people like pineapple on their pizza**), and we determine whether the sequence's language is English (E), Polish (P) or Spanish (S). We will treat each character as a separate observation. We only know the probabilities of vowels **a**, **e**, **i**, **o**, **u**, **y**, so we will treat all other characters as one class. The numbers are taken from `https://en.wikipedia.org/wiki/Letter_frequency#Relative_frequencies_of_letters_in_other_languages`.

| $p(\downarrow \mid \rightarrow)$ | E | P | S |
|---|---|---|---|
| a | 0.08167 | 0.10503 | 0.11525 |
| e | 0.12702 | 0.07352 | 0.12181 |
| i | 0.06966 | 0.08328 | 0.06247 |
| o | 0.07507 | 0.02445 | 0.08683 |
| u | 0.02758 | 0.02062 | 0.02927 |
| y | 0.01974 | 0.03206 | 0.01008 |
| other | 0.59926 | 0.66104 | 0.57429 |

Implement the language classifier and answer the following:

(a) **[0.5p]** We can implement the Naive Bayes classifier either by multiplying probabilities or by adding log-probabilities. Which one is better and why?

(b) **[1.5p]** Assuming equal class probabilities $p(C = E) = p(C = P) = p(C = E) = 1/3$, what is the language of the following phrases, according to the classifier?

   i. bull,
   ii. burro,
   iii. kurczak,
   iv. pollo,
   v. litwo, ojczyzno moja, ty jesteś jak zdrowie,
   vi. dinero,
   vii. mama just killed a man put a gun against his head,
   viii. maradona es mas grande que pele.

(c) **[0.5p]** Let us assume that the a priori probabilities are $p(E) = 0.5$, $p(P) = 0.2$, $p(S) = 0.3$. How will the results change?

(d) **[0.5p]** Try to fool the classifier! Create a phrase which is misclassified.

(e) **[0-1p bonus]** What happens when a Naive Bayes classifier is applied to a document with out-of-vocabulary words? Propose some solutions. Relate them to the concept of Bayesian priors discussed during the lecture.

## 3 Maximum likelihood estimation

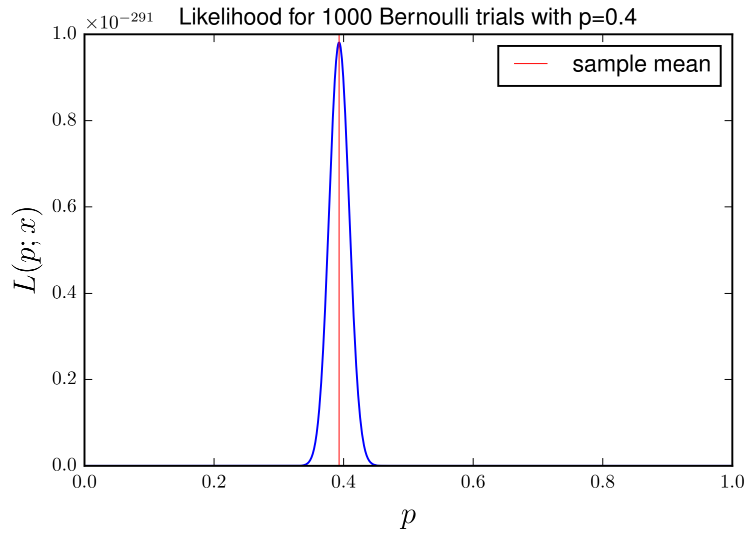
**Problem 5.** **[3p]** Given observations $x_1, \ldots, x_n$ coming from a certain distribution, prove that MLE of a particular parameter of that distribution is equal to the sample mean $\frac{1}{n}\sum_{i=1}^{n} x_i$:

(a) **[1p]** Bernoulli distribution with success probability $p$ and MLE $\hat{p}$,

(b) **[1p]** Gaussian distribution $\mathcal{N}(\mu, \sigma)$ and MLE $\hat{\mu}$,

(c) **[1p]** Poisson distribution $Pois(\lambda)$ and MLE $\hat{\lambda}$.

**Problem 6.** **[1p]** Draw $n$ random samples from the Bernoulli distribution. Plot sample mean and likelihood as a function of $p$. Experiment with $n \in \{50, 100, 500, 1000\}$ and $p \in \{0.4, 0.5\}$. You should obtain a plot like the one below:



**Tip:** To make a plot, evaluate $L(p; x)$ in a finite set of points, for instance:

```
p = linspace(0.0, 1.0, 1000)
```

**Problem 7.** **[2p]** There is a game called **rock, paper, scissors**. Klapaucius has built a robot which plays the game (it displays **R**, **P** or **S** – the first letter of the appropriate symbol – on a monitor). The robot has won the world championship, however, its memory drive – including its configuration – has been irreparably damaged due to an overly excessive celebration, including the consumption of unhealthy amounts of grease.

The robot followed a rather simple algorithm. The choice of the symbol was randomized, and it depended only on the previous symbol. The algorithm has nine parameters – $\{p(a|b) : a, b \in \{R, P, S\}\}$ – denoting the probability of showing symbol $a$ if the most recently shown symbol was $b$. Klapaucius has a terrible memory for numbers and he does not remember the configuration. Fortunately, he has got hold of the video footage from which he was able to decipher the sequence of symbols shown. The sequence is as follows:

PPRSSRSPPRSPRRSPPPSSPRSPSPSRSP

(a) **[1p]** Find the most likely configuration, given that the first symbol has been sampled from a uniform distribution.

(b) **[1p]** Trurl is the runner-up. He decided to copy Klapaucius's robot. He decided, however, that there is no need for the probabilities to be based on the previous symbol. His configuration had three parameters – $p(R)$, $p(P)$, $p(S)$. He set them according to the same footage. Find the parameters. Simulate a million rounds between the two robots. Which robot wins more frequently?