Exercise 1 – Probability & Python

1. There are 1000 coins, which look identical. However, 999 of them are “fair” (i.e. when tossing the coin, the probability to get “heads” is 0.5), and one coin is forged (i.e. when tossing the coin, the probability to get “heads” is 0.8). Assuming we selected on random one coin and toss it 10 times, and got “heads” in 7 out of the 10 tosses. What is the probability that this is the forged coin?

2. In a faraway country, families favor girls over boys. So they settle for one male child. Thus, each family gives birth until their first son is born, and then they stop (they never take the risk of having two boys – double trouble...). As a result, families in this country are of the following types:
   a. Boy
   b. Girl, Boy
   c. Girl, Girl, Boy
   d. Girl, Girl, Girl, Boy
   e. ...

   Are there more girls or more boys in this country?
   You may assume that $Pr(boy) = Pr(girl) = 0.5$

3. Given a 2-dimensional random variable $\boldsymbol{x} \sim N_2(\boldsymbol{\mu}, \Sigma)$

$$\Pr(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

   Where $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

   <u>Show that</u>

   a. The marginal distributions are normal, such that

$$x_i \sim N(\mu_i, \sigma_i^2), \quad i = 1,2$$

   <u>Hints</u>

   - Consider using the explicit form of the bivariate Normal distribution
     - $f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$
   - Recall that $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$
     Before solving the integral, consider changing variable $z_2 = \frac{x_2-\mu_2}{\sigma_2}$ and complete to square in z_2
   - The following integrals may be useful (which one to use depends on your specific solution)
     - $\int_{-\infty}^{\infty} \exp[-a(x+u)^2] dx = \sqrt{\pi/a}$
     - $\int_{-\infty}^{\infty} \exp[-(ax^2+bx+c)] dx = \sqrt{\pi/a} \exp\left(\frac{b^2}{4a} - c\right), \quad a > 0$

   b. The conditional distribution is normal, such that

$$x_1|x_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}, \sigma_1^2(1-\rho^2)\right)$$

Hints

- Recall that $f_{X_1|X_2=x_2}(x_1) = \frac{f(x_1,x_2)}{f_{X_2}(x_2)}$

  Get to the form $(A^2 - 2AB + B^2) = (A-B)^2$ in the exponent
- Reshape the result such that it will resemble the form of a 1-dim Normal density, and identify the mean and the variance

4. Pearson's Correlation Coefficient is defined as follows

$$\rho = \frac{cov(X,Y)}{\sigma_x\sigma_y} = \frac{E\left[(x-\mu_x)(y-\mu_y)\right]}{\sqrt{E[(x-\mu_x\ )^2]}\sqrt{E[(y-\mu_y)^2]}}$$

Show that $-1 \le \rho \le 1$ (Hint: use Cauchy–Schwarz inequality)

5. We toss 1000 <u>identical</u> coins, each coin 10 times. The probability of each coin showing "heads" is $p \in (0,1)$. Let θ_i be the <u>number</u> of times the i-th coin showed "heads".
   a. What is the distribution of each θ_i?
   b. What is the mean $\mu = \mathbb{E}[\theta_i]$?
   c. We would like to use $\hat{\theta}$ as an estimate for μ. Let us analyze the error of this estimate.
      Use Hoeffding's inequality to compute the smallest error margin possible with confidence 0.95. That is, find the smallest ϵ that holds $\Pr\left[\left|\hat{\theta}-\mu\right| > \epsilon\right] \le 0.05$

Hints:

Recall Hoeffding's inequality:

Let $\theta_1, \dots, \theta_n$ be a sequence of i.i.d. random variables.

Assume all variables have the same mean, i.e. $\forall i\colon \mathbb{E}[\theta_i] = \mu$, and are bounded s.t.

$$\Pr[a \le \theta_i \le b] = 1.$$

Then, for any $\epsilon > 0$

$$\Pr\left[\left|\hat{\theta}-\mu\right| > \epsilon\right] \le 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}$$

where $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta_i$ is the empirical mean.

6. Let A_1, A_2 be events in a probability space.

   Prove that:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) \le \sum_{i=1}^{n} P(A_i)$$

Computer Assignment – Numpy and SciPy

Please go over the document *'A Short Introduction to Python – Numpy, Scipy'* located in the course's website. **This will greatly help you in the upcoming exercises.**
Numpy is the standard that is used for machine learning in Python. It allows you to write code that is closer to your math very efficiently. This will require you at times to write matrix multiplication rather than loops.

Attached please find numpy_exercise.py containing

- An assignment – the description is documented in the file
- A template code that will get you going

You should implement your code **without loops**, please notice that each function can be implemented with two lines of code or less.
Test your code and submit your code in the same file.

Computer Assignment – Learn Pandas, Matplotlib and Scikit-learn

Please open the iPython notebook "Introduction_to_Pandas,_Matplotlib_and_Scikit_learn.ipynb" and go over it. It will demonstrate the basics of working with Pandas, Matplotlib and Scikit-learn, libraries that are going to be used throughout the course in the exercises (and are widely used in the academia and industry). At the end of the script there are several simple assignments.

How to open an iPython notebook file?

We recommend the following options:

1. Using anaconda's Jupyter lab.
2. Using Colaboratory. A Google drive extension that allows you to run notebooks from your browser.