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 0.9). Assuming we selected on random one coin and toss it 10 times, and got “heads” in 9 out of the 10 tosses. What is the probability that this is the forged coin?

Recall Bayes formula

\[
Pr(B_i|A) = \frac{Pr(A|B_i) Pr(B_i)}{\sum_{j=1}^{k} Pr(B_j) Pr(A|B_j)} = \frac{Pr(A|B_i) Pr(B_i)}{Pr(A)}
\]

Where \(B_1, ..., B_k\) are mutually exclusive and form a partition of the sample space

For this exercise, the Bayes formula should be interpreted as follows

\[
\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}
\]

Denote: \(A\) - the event of having 9 “heads” and 1 “tail”, \(B_L\) - a Legal coin, \(B_F\) - a Forged coin.

Assigning the values given in the exercises, we get

\[
Pr(B_F|A) = \frac{Pr(A|B_F) Pr(B_F)}{Pr(A|B_F) Pr(B_F) + Pr(A|B_L) Pr(B_L)} = \frac{0.9^9 \times 0.1 \times 0.001}{0.00101432} \cong 0.038
\]

2. In a faraway country, they have a strange culture where they favor girls over boys. Additionally, they know that boys are trouble, so they settle for one in a family. Thus, each family give 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?

Average number of boys = Average number of girls.
This is true in two possible scenarios
1) There is an upper bound to the permissible number of births for each number
2) There is NO limit to number of births for each number

The argument makes use of the fact that per birth, the probabilities to have a boy or a girl are equal. Hence, in the first birth of all women in the country, the average
The number of boys is equal to the average number of girls. This is also true for the birth of the second child, etc. Please note that whoever calculated the sums of sequences, it should be noted that if there is an upper bound on the number of births (finite sequence) – the last birth is not necessarily a boy.

3. Given a random sample \( \{x_1, x_2, ..., x_N\} \) Calculate the Maximum Likelihood Estimator (MLE) for the parameters \( \Theta \) of the following
   a. Binomial – \( \Pr(k|n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \)
      
      \[
      L({x_1, x_2, ..., x_N}|p) = \prod_{i=1}^{N} \left( \frac{n}{x_i} \right) p^{x_i} (1 - p)^{n-x_i}
      \]
      
      \[
      \log L({x_1, x_2, ..., x_N}|p) = \sum_{i=1}^{N} \left( \frac{n}{x_i} \right) + \sum_{i=1}^{N} x_i \log p + \sum_{i=1}^{N} (n - x_i) \log(1 - p)
      \]
      
      \[
      \frac{\partial \log L}{\partial p} = 0 \quad \Rightarrow \quad \frac{\sum_{i=1}^{N} x_i}{p} = \frac{\sum_{i=1}^{N} (n-x_i)}{1-p}
      \]
      
      \[
      p(\sum_{i=1}^{N} n - x_i) = (1 - p) \sum_{i=1}^{N} x_i
      \]
      
      \[
      \hat{p} = \frac{1}{nN} \sum_{i=1}^{N} x_i
      \]
   
   b. Normal – \( \Pr(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \)
      
      \[
      \log L({x_1, x_2, ..., x_N}|\mu, \sigma^2) = -n \left( \log \pi + \frac{1}{2} \log \sigma^2 \right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
      \]
      
      \[
      \frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i
      \]
      
      \[
      \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2
      \]
   
   c. Poisson – \( \Pr(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \)
      
      \[
      \log L({k_1, k_2, ..., k_n}|\lambda) = \sum_{i=1}^{n} k_i \log \lambda - \sum_{i=1}^{n} k_i! - \sum_{i=1}^{n} \lambda
      \]
      
      \[
      \frac{\partial \log L}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} k_i - \sum_{i=1}^{n} \lambda = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} k_i
      \]
4. Pearson’s Correlation Coefficient is defined as follows
\[ \rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sqrt{E[(x - \mu_x)^2]} \sqrt{E[(y - \mu_y)^2]}} \]
Show that \(-1 \leq \rho \leq 1\) (Hint: use Cauchy–Schwarz inequality)

Reminder: Cauchy–Schwarz inequality
\[ |<x, y>|^2 \leq <x, x><y, y> \]
Equality holds iff \(x\) and \(y\) are linearly dependent

Denote \(<X, Y> \equiv E(X, Y)\), then by Cauchy–Schwarz inequality,
\[ |E(X, Y)|^2 \leq E(X^2)E(Y^2) \]

And
\[ |\text{Cov}(X, Y)|^2 = \left| E\left( (X - \mu_x)(Y - \mu_y) \right) \right|^2 = \left| <X - \mu_x, Y - \mu_y> \right|^2 \]
\[ \leq <X - \mu_x, X - \mu_x><Y - \mu_y, Y - \mu_y> \]
\[ = E((X - \mu_x)^2) E\left((Y - \mu_y)^2\right) \]
\[ = \text{Var}(X) \text{Var}(Y) \]