DEPARTAMENTO DE ELETRÓNICA, TELECOMUNICAÇÕES E INFORMÁTICA

LICENCIATURA EM ENGENHARIA DE COMPUTADORES E INFORMÁTICA

ANO 2025/2026

MÉTODOS PROBABILÍSTICOS PARA ENGENHARIA DE COMPUTADORES E INFORMÁTICA

PRACTICAL GUIDE NO. 1

Task 1 - Simulation of simple random experiments with Matlab

1. Consider the random experiment of tossing a balanced coin 3 times. The aim is to estimate by simulation the probability of getting 2 heads after the 3 tosses.

To estimate the probability by simulation, one needs to perform the random experiment several times and calculate the relative frequency of the result on interest. In Matlab, one possible way to implement this simulator is as follows (assuming the experiment is run 10000 times):

```
%% Code 1
% Generate a matrix with 3 rows and 10,000 columns of random numbers
% between 0.0 and 1.0 (i.e., each column represents an experiment):
experiments = rand(3,10000);
% Generate a matrix with 3 rows and 10,000 columns with the value 1 if
% the value of the previous matrix is less than 0.5 (i.e., if it came
% up heads) or with the value 0 otherwise (i.e., if it came up tails):
tosses = experiments < 0.5; % 0.5 corresponds to the prob. of heads
% Generate a row vector with 10,000 elements with the sum of the values
%of each column of the previous matrix (i.e., the number of heads in
% each experiment):
results = sum(tosses);
% Generate a row vector with 10,000 elements with the value 1 when the
% value of the previous vector is 2 (i.e., if the experiment gave 2 heads)
% or 0 when it is different from 2:
successes = results == 2;
% Determine the result by dividing the number of experiments with 2 heads
% by the total number of experiments:
probSimulation = sum(successes)/10000
```

The proposed code is developed step by step for easier understanding. Individual operations can be combined to avoid intermediate matrices and make code execution more efficient. Furthermore, it is useful to use initial variables for the problem parameters to make it easier to adapt the script to other cases of interest. For example, another alternative to implement the same simulator is:

```
%% Code 1 - second version
N = 1e4; % number of experiments
p = 0.5; % probability of heads
k = 2; % number of heads
n = 3; % number of tosses
tosses = rand(n,N) < p;
successes = sum(tosses) == k;
probSimulation = sum(successes)/N
```

(a) Analyze and test the 2 versions of the Matlab code provided.

(b) Estimate, using the 2 versions of the code, the probability of getting 2 heads in 3 tosses of a balanced coin (run the simulation several times and take conclusions).

2. Estimate:

(a) the probability of getting 6 heads in 15 tosses of a balanced coin,

(b) the probability of getting at least 6 heads in 15 tosses of a balanced coin.

3. To easy the calculation of other cases like those discussed in the previous points, create a Matlab function that allows to estimate the probability through simulation. The function should have the following input parameters: ***p, number of tosses, number of desired heads, and number of experiments to be performed***. You should use appropriate names for the function and the input parameters.

(a) Re-estimate the probabilities of the previous questions using the function developed.

(b) Estimate the probabilities of the entire sampling spaces for the following numbers of tosses: 20, 40 and 100. Make 3 graphs, using the Matlab *stem* function, of the estimated probabilities for the three number of tosses.

4. Now, the aim is to calculate analytically the probabilities estimated in the previous exercises.

This probability is given by the following expression:

$$P(k) = C_k^n p^k (1-p)^{n-k}$$

where $\boldsymbol{p}$ is the probability of the event (for example: if the event is "heads" on each toss, p = 0.5), $\boldsymbol{k}$ is the number of events that occurred in $\boldsymbol{n}$ repetitions of the random experiment.

In Matlab, this expression is determined as follows:

```
%% Code 2 - Analytical calculation of probability in series of
% Bernoulli experiments
% Data related to problem 1
p = 0.5;
k = 2;
n = 3;
prob= nchoosek(n,k)*p^k*(1-p)^(n-k); % nchoosek(n,k)= n!/(n-k)!/k!
```

Calculate the analytical value for each of the problems in questions **1** and **2**, and compare the results obtained with the estimated values. (What do you conclude?)

5. On a given data link, the BER (bit error rate) is 10^{-5}, and the errors in the different bits of a data packet are statistically independent. Determine theoretically:

(a) the probability of a data packet of size 100 Bytes to be received without errors,

(b) the probability of a data packet of size 1000 Bytes to be received with at least 2 errors.

6. (Optional) Consider a factory production process that produces faucets in which the probability of each faucet being produced defective is 10%. At the quality control process, a sample of 5 faucets is selected.

(a) Calculate analytically and by simulation the probability that 3 faucets in the sample are defective.

(b) Calculate analytically and by simulation the probability that at most 2 of the sample faucets are defective.

(c) Based on simulation, compute in Matlab the histogram representing the probability distribution of the number of defective faucets in the sample.

Task 2 - Probability, conditional probability, and independence

Answer the following questions using simulations in Matlab and/or analytically and, whenever asked, compare both results.

1. Consider families with children in which the probability of having boys is equal to the probability of having girls.

(a) Obtain, through simulation, an estimate of the probability of the event "having at least one son" in families with two children.

(b) Determine the theoretical value of the event in the previous paragraph and compare it with the estimated value obtained through simulation. Do you obtain the same values? Why?

(c) Assume that for a family with two children chosen at random, we know that one of the children is a boy. What is the probability that the other child will also be a boy? Determine the theoretical value of this probability and estimate the same probability by simulation.

(d) Knowing that the first child in a family with two children is a boy, determine by simulation the probability of the second child being a boy. What can be concluded from the result obtained regarding the independence of events?

(e) Consider a family with five children. Knowing that at least one of the children is a boy, obtain by simulation an estimate of the probability that one of the others (and only one) is also a boy.

(f) Repeat point (e) but estimating the probability that at least one of the others is also a boy.

2. Consider the following "game": blindfolded throwing of $\boldsymbol{n}$ darts, one at a time, at $\boldsymbol{m}$ targets, ensuring that each dart always hits a target (and only 1).

(a) Estimate by simulation the probability that no target is hit more than once when $\boldsymbol{n} = \mathbf{20}$ darts and $\boldsymbol{m} = \mathbf{100}$ targets.

(b) Estimate by simulation the probability that at least 1 target is hit two or more times when $\boldsymbol{n} = \mathbf{20}$ darts and $\boldsymbol{m} = \mathbf{100}$ targets.

(c) Consider the values of $\boldsymbol{m} = \mathbf{1000}$ and $\boldsymbol{m} = \mathbf{100000}$ targets. For each of these values, run the necessary simulations to draw a graph (using Matlab *plot* function) of the probability of item (b) as a function of the number of darts $\boldsymbol{n}$ (consider **n** from 10 to 100 in increments of 10). The two graphs should be subgraphs of the same figure (use Matlab *subplot* instruction). Compare the results of the two cases and draw conclusions.

3. (Optional) Consider an array of size $\boldsymbol{T}$ that serves as the basis for the implementation of an associative memory (for example in Java). Assume that the hash function returns a value between 0 and $\boldsymbol{T} - 1$ with all values equally likely.

(a) Determine by simulation the probability of at least one collision (at least 2 keys mapped by the hash function to the same array position) if 10 keys are inserted into an array of size $\boldsymbol{T} = \mathbf{1000}$.

(b) Plot the probability of (a) as a function of the number of keys for all relevant values in an array of size $\boldsymbol{T} = \mathbf{1000}$.

(c) For 50 keys, plot the probability (estimated by simulation) of no collisions as a function of the size $\boldsymbol{T}$ of the array (consider $\boldsymbol{T}$ sizes from 100 to 1000 in increments of 100).

4. Consider a party attended by ***n*** persons.

(a) Determine by simulation the smallest value of ***n*** for which the probability of two or more persons having the same birthday (month and day) is greater than 0.5 (assume a year with 365 days)?

(b) What must be the smallest value of ***n*** for the probability in the previous item to become greater than 0.9?

5. Consider a six-sided die numbered 1 to 6 rolled twice. Assume that the die is balanced (equal probability for all faces to land on top). Consider the following events:

"A – the sum of the two values is equal to 9";

"B – the second value is even",

"C – at least one of the values is equal to 5", and

"D – none of the values is equal to 1".

(a) Estimate by simulation the probability of each of the four events.

(b) Determine theoretically whether events A and B are independent.

(c) Determine theoretically whether events C and D are independent.

6. Consider a wireless link between two hosts where the probability of the transmitted data packets being received with errors is 0.1% in normal link conditions or 10% with external interferences. The probability of the link being with external interferences is 2%. In reception, the hosts are able detect if each data packet is or is not received with errors.

(a) Compute analytically the probability of a data packet being received with errors.

(b) When the data packet is received with errors, compute analytically the probability of the link being in normal conditions and the probability of the link being with external interference.

7. (Optional) Consider a company with 3 programmers (André, Bruno, and Carlos). The probability of a program from each of them having errors ("bugs") and the number of programs developed by each programmer are the values presented in the following table.

Programmer	Error (“bug in a program”)	Number of programs
André	0.01	20
Bruno	0.05	30
Carlos	0.001	50

The company CEO randomly selects one of the 100 programs developed by the 3 programmers and discovers that it contains an error. Compute analytically:

(a) the probability that the selected program is of Carlos,

(b) the most likely programmer of the program selected by the CEO.