Distributions of random variables. Histograms

Laboratory 1, DEPI

Objective

Gain familiarity with the Matlab environment, and with the functions useful for working with random variables.

Theoretical aspects

The following aspects shall be explained.

1. Random variables in Matlab
    1. Normal distribution
        - generating one value, or a vector of values from the standard normal distribution, using `randn()`
        - generating values for a specified average and deviation value
    2. Uniform distribution
        - generating one value, or a vector of values from the uniform distribution, using `rand()`
        - generating values for specified interval of the uniform distribution
2. Functions in Matlab
    - defined in separate file
    - definition of a function
3. Histogram plots
    - what are they
    - the Matlab `hist()` function

Exercises

1. Generate a vector with 1000 values from the normal distribution $\mathcal{N}(2,2)$ and plot the values
2. Generate a vector with 1000 values from the uniform distribution $\mathcal{U}[-4,10]$ and plot the values
3. Create a Matlab function `myCDF()` that estimates the cumulative distribution function (CDF) from a vector of data
   - the function requires two arguments and returns one value `p = myCDF(v,x)`
   - `v` is a vector, `x` and `p` are scalar numbers
   - the function computes how many elements from `v` are smaller or equal than `x`, divided to the total number of elements of `v`
   - given a long sequence `v`, what does the return value `p` mean?
4. Use the `myCDF()` function to compute the CDF of the normal / uniform distributions from the two vectors generated previously.
   - generate a vector n of 50 values uniformly spread between -8 to 18
   - apply `myCDF()` to all values from the vector n, for the two vectors generated previously
   - plot the results of the function against the values of n
5. Plot the probability density function from a vector of data, based on the `myCDF()` function defined above
   - generate a vector n of 50 values uniformly spread between -8 to 18
   - apply `myCDF()` to all values from the vector n, for the two vectors generated previously
   - compute the difference between every two adjacent elements of the resulting vector
   - plot the results of the function against the values of n
6. Create another Matlab function `myPDF()` that directly estimates the probability density function from a vector of data
   - the function requires three arguments and returns one value `p = myPDF(v,x,epsilon)`
   - `v` is a vector, `x`, `epsilon` and `p` are scalar numbers
   - the function computes how many elements from `v` are in the interval $[x-\epsilon, x+\epsilon]$, divided to the total number of elements of `v`, and also divided to `epsilon`
   - given a long sequence `v`, what does the return value `p` mean?
7. Plot the probability density function estimated from a vector of data
   - generate a vector n of 50 values uniformly spread between -8 to 18

- apply `myPDF()` to all values from the vector n, for the two vectors generated previously (use `epsilon = 0.3`
- plot the results of the function against the values of n

8. Plot the histograms of the two data vectors with `hist()`

Final questions

1. What should we modify in order to get more accurate results / plots of these functions?