Report lab L3

Alessandro Mosca

November 9, 2023

1 Problem Proposition

In this Laboratory we were asked to generate some random variables, in particular we were asked to generate these different distribution:

- Rayleigh(σ):
$$f(x:\sigma) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$$
- Lognormal(μ,σ^2):
$$f(x:\mu,\sigma^2) = \frac{e^{-\frac{((\ln(x-\mu)^2}{2\sigma^2})}}{x\sqrt{2\pi}\sigma}$$
- Beta(α,β):
$$f(x:\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \qquad \alpha > 1, \beta > 1$$
- Chisquare(n):
$$f(x:n) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}x^{\frac{k}{2}-1}e^{-\frac{x}{2}} \qquad n \geq 1$$
- Rice(ν,σ)
$$f(x:\nu,\sigma) = \frac{x}{\sigma^2}e^{-\frac{x^2+\nu^2}{2\sigma^2}}I_0(\frac{x\nu}{\sigma^2}) \qquad \nu \geq 0, \sigma \geq 0$$

To do this we used the methods proposed during class, we then want to compare the empirical results with the arithmetically computed values of the first two moments considering different number of samples and the respective CDF and PDF.

2 Methods used

In order to generate the random variables the methods used are:

- Inverse-Transform technique:
  We compute the inverse of the pdf and then use a number generated from a uniform random variable to compute the random variable.
  We used this method for:
  - Rayleigh distribution
- Acceptance/Rejection technique:
  We generate two numbers from two uniform random variables and use it as coordinates of a point in a 2D plane, we can then test if the generated point is under the pdf, where we accept it or over it where we discard it.
  We used this method for:
  - Lognormal distribution
  - Beta distribution

Random Variable	Number of samples	Empirical Mean	Expected Mean	Relative error
Rayleigh(1)	1000	1.2472321934294315	1.2533141373155001	0.4852689126363551 %
Rayleigh(1)	10000	1.2401876674737056	1.2533141373155001	1.0473407624612305 %
Rayleigh(1)	100000	1.2494088537095283	1.2533141373155001	0.31159654947614923 %
Lognormal(0,0.1)	1000	1.0454614519298457	1.0512710963760241	0.11218489712871782 %
Lognormal(0,0.1)	10000	1.0571711225070626	1.0512710963760241	0.5612278461166889 %
Lognormal(0,0.1)	100000	1.0500917289780107	1.0512710963760241	0.5526304743092024 %
Beta(1,1)	1000	0.49460323202770756	0.5	1.0793535944584876 %
Beta(1,1)	10000	0.5010967630900174	0.5	0.2193526180034766 %
Beta(1,1)	100000	0.5005119437745176	0.5	0.10238875490351518 %
Chisquare(3)	1000	3.1094344379847088	3	3.647814599490292 %
Chisquare(3)	10000	2.985239038081668	3	0.4920320639443994 %
Chisquare(3)	100000	2.9966231859430925	3	0.11256046856358282 %
Rice(4,1)	1000	4.08230802316155	4.127193542536757	1.0875554759571087 %
Rice(4,1)	10000	4.114785463542407	4.127193542536757	0.30064204323027244 %
Rice(4,1)	100000	4.131589559740467	4.127193542536757	0.10651347358448117 %

Table 1: Table containing the results for the mean.

- Convolution Method:
  We can use this method only for random variables that are sum of other random variables, in this case we can generate the variables independently and then compute the value of the random number.
  We used this method for:
  - Chisquare distribution
  - Rice distribution

3 Results

After running the simulation we can show the results in table 1 where we have the results for the mean and in table 2 where we have the results for the variance.
The mean has been computed using the average function in numpy and the variance has been computed by squaring the result of the std function in numpy, both applied after the generation of the specified number of samples.
We can see that the results approximate the mean and the variance better and better as the number of samples grows, this happens because the arithmetic value for mean and variance are computed as if the distribution had an infinite number of samples and so, by generating more and more numbers, we are getting closer and closer to the expected value.
It might also happen that a lower number of samples actually gives a result closer to the expected value, this can be explained because, due to the stochasticity of the problem, it might happen that the generated number are more representative of the distribution.

4 Plotting the PDF and CDF

After analyzing the mean and the variance we want to compare both the PDF and the CDF of the distribution with the empirically computed ones, in this case, we want to show the results for the Rice distribution considering as parameters

- $\nu = 4$
- $\sigma = 1$

Random Variable	Number of samples	Empirical Variance	Expected Variance	Relative error
Rayleigh(1)	1000	0.45369832909074215	0.42920367320510344	5.707000525583445 %
Rayleigh(1)	10000	0.4413222591401149	0.42920367320510344	2.8235047115312937%
Rayleigh(1)	100000	0.4300655244634388	0.42920367320510344	0.20080239572496042 %
Lognormal(0,0.1)	1000	0.11878033932978287	0.11623184008452231	2.1925999307997888 %
Lognormal(0,0.1)	10000	0.11597022541127343	0.11623184008452231	0.22508004094113776 %
Lognormal(0,0.1)	100000	0.11756596546894622	0.11623184008452231	1.1478140442874827 %
Beta(1,1)	1000	0.08897995940463077	0.08333333333333333	6.77595128555693 %
Beta(1,1)	10000	0.0833515836761145	0.08333333333333333	0.021900411337411052 %
Beta(1,1)	100000	0.08324168779317467	0.08333333333333333	0.10997464819038627 %
Chisquare(3)	1000	5.188396137200207	6	13.52673104666322 %
Chisquare(3)	10000	6.219369978466733	6	3.6561663077788817 %
Chisquare(3)	100000	5.975219283188508	6	0.4130119468581957 %
Rice(4,1)	1000	0.9726757319786243	0.9662734624428921	0.6625732553542604 %
Rice(4,1)	10000	0.9837551236644062	0.9662734624428921	1.809183621509967 %
Rice(4,1)	100000	0.9682160195171323	0.9662734624428921	0.2010359540796132 %

Table 2: Table containing the results for the mean.

To compute the empirical PDF we use the histogram function in numpy, this returns the value of the histogram and the number of generated bins, we then take these values and compute the empirically discovered PDF and CDF.
In order for the results to be correctly evaluated we used the scipy.stats library to generate the analytical PDF and CDF and plotted the results using matplotlib.
As we can see in Fig.1 with a low number of samples the empirical PDF is far from the analytical one, but looking at the results in Fig.2 and Fig.3 we can see the results get closer and closer to the expected output as the number of samples grows. This happens because the analytical result behaves as if it had an infinite number of samples and generating more numbers makes us closer and closer to the expected result.
The same can be seen for the empirical CDF

5 Conclusions

In this laboratory we were able to show how we can generate non-standard distributions using different algotrithm and that by generating a large number of samples the mean, variance, PDF and CDF of the distribution get closer and closer to the expected values

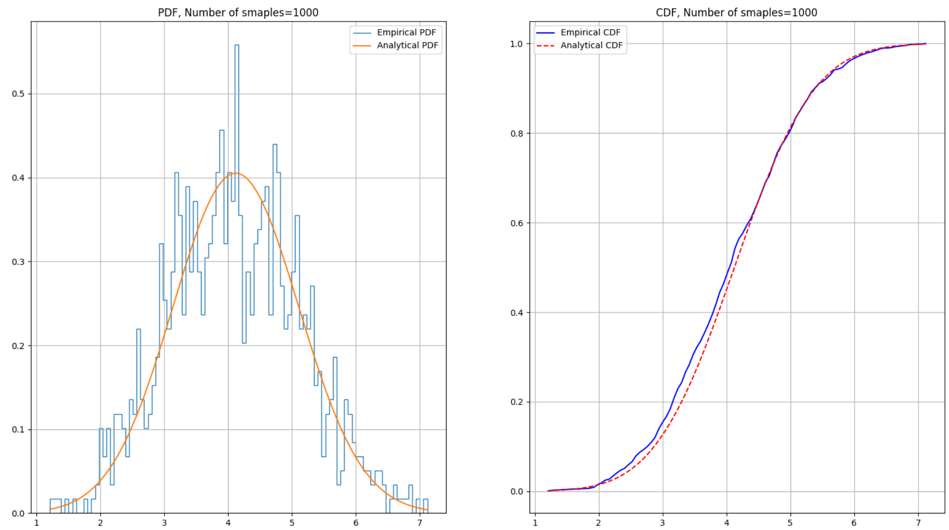

Figure 1: Results with 1000 samples

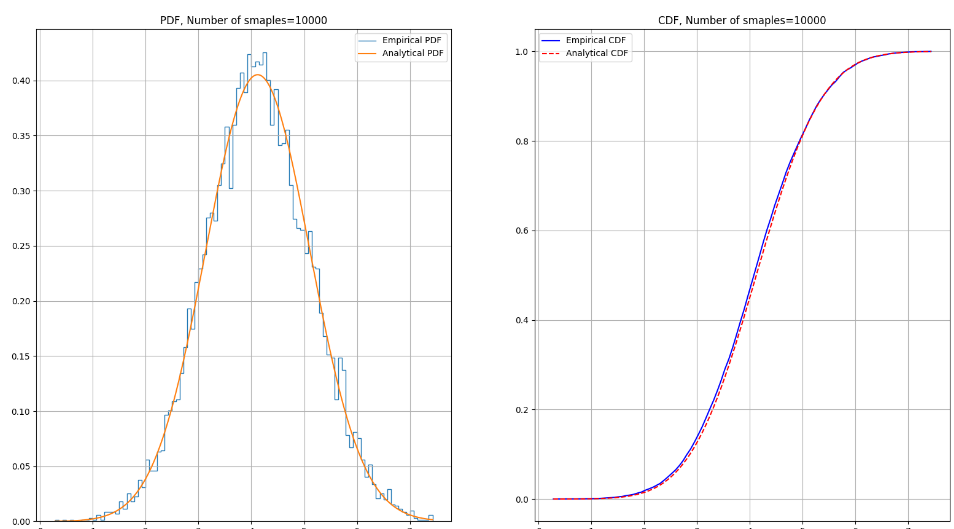

Figure 2: Results with 10000 samples

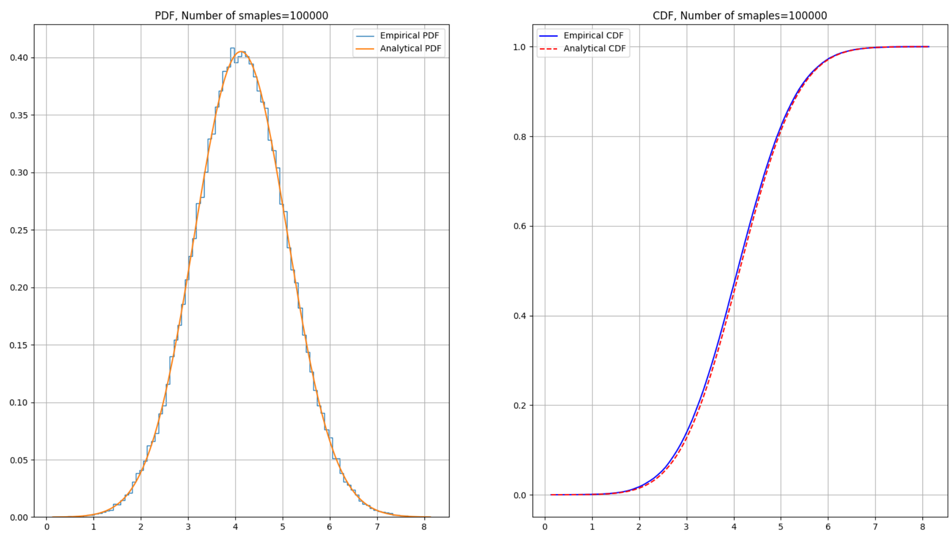

Figure 3: Results with 100000 samples