Lab L3: Generation of Random Variables

Emanuele Pietropaolo
Politecnico di Torino
Student id: s319501
emanuele.pietropaolo@studenti.polito.it

***Abstract*—The aim of this Lab is to develop functions that can generate random variables that follow specified distributions derived from a uniform distribution. Various techniques will be used for this task, ranging from the use of the inverse cumulative method to the implementation of a custom method tailored to the target distribution.**

I. Problem overview

Generating random variables that follow desired distributions is a fundamental task in stochastic computing simulation. Computers can use some specific algorithms to generate uniform random variables that, given the state of memory, produce a sequence of numbers that appear random; these numbers are called **Pseudorandom numbers**. The uniform distribution is typically the default generator in most programming languages due to its simplicity and efficiency. For this reason, being able to generate other distributions from the uniform one is an important task and can be done in several ways. These include the inverse cumulative method and ad-hoc methods that are specifically designed for the target distribution. The aim is to develop a generator class capable of generating variables that follow these distributions: Rayleigh distribution, LogNormal distribution, Beta distribution, Chi-Square distribution and Rice distribution. Each method takes as input the parameters that describe the target distribution that the data should follow.

II. Proposed approach

A. Rayleigh distribution

The Rayleigh distribution is a continuous probability distribution for non-negative-valued random variables. This distribution is described by its input parameter σ. Random variables that follow this distribution can be generated using the inverse cumulative method. Knowing that its *Probability Density Function* is defined as

$$f(x;\sigma) = \begin{cases} \frac{x}{\sigma^2} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right), & \text{for } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

and that its *Cumulative distribution function* is given as:

$$F(x;\sigma) = \begin{cases} 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right), & \text{for } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

we can generate positive Rayleigh random variables ($X \sim Rayleigh(\sigma)$) from the inverse of the CDF:

$$X = \sigma \cdot \sqrt{\ln(1-U)}$$

where $U \sim Uniform(0,1)$ is a uniform distributed random variable.

B. LogNormal distribution

The log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. As the definition suggests, to generate a log-normal random variable we can exponentially elevate a normal random variable. If the random variable Y is normally distributed ($Y \sim N(\mu, \sigma^2)$), then $X = exp(Y)$ has a log-normal distribution ($X \sim Lognormal(\mu, \sigma^2)$).

C. Beta distribution

The beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ in terms of two positive parameters (α,β) that control the shape of the distribution. I generated beta random variables derived from two gamma distribution. In fact, if $X \sim Gamma(\alpha, \theta)$ and $Y \sim Gamma(\beta, \theta)$ are independent gamma variate variables, then:

$$\tfrac{X}{X+Y} \sim Beta(\alpha, \beta)$$

D. Chi Squared distribution

The chi-squared distribution with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. I used two different method for generating chi-squared random variables:

1) The inverse cumulative method for the case $k = 2$ in which the chi-squared distribution can be described by an exponential with mean 2
$$X \sim \chi^2_2 \text{ then } X \sim Exp(1/2)$$
2) As the definition suggests, from the sum of the squares of k independent standard normal random variables
$$X = \textstyle\sum_1^n N^2$$
where $N \sim N(0, 1)$.

E. Rice distribution

The Rice distribution is the probability distribution of the magnitude of a circularly-simmetric bivariate normal random variable. It's described by two input parameters: v and σ. Rice random variables can be generated from independent normal random variables. In fact, if $X \sim N(vcos\theta, \sigma^2)$ and $Y \sim N(vsin\theta, \sigma^2)$ then:

$$R = \sqrt{X^2 + y^2} \sim Rice(|v|, \sigma)$$

where θ is any real number. I chose θ to obtain $cos\theta = 1$ and $sin\theta = 0$.

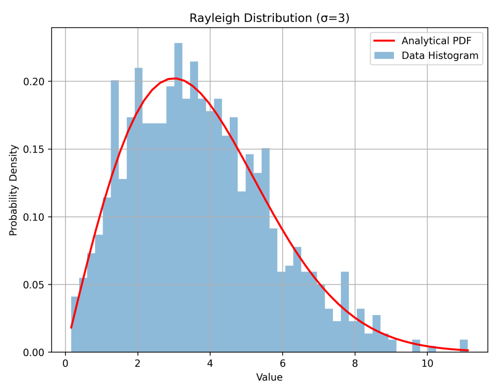
III. Results

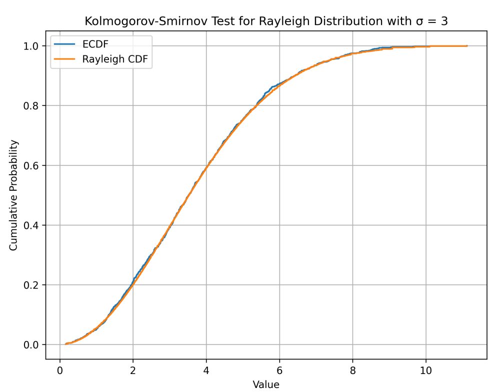
I employed three approaches to evaluate the data's goodness of fit.

- **Analizing the first two momentums**: the analytical first two momentums can be compared with the empirical one and the comparison can give useful insight for studying if the data follow the desired distribution.
- **Chi-squared test**: is a statistical hypothesis test used in the analysis of contingency tables. It examines whether two variables follow the same distribution. I compare the empirical pdf with the analytical one.
- **Kolmogorov-Smirnov test (KS-test)**: also this test compare if two variables follow the same distribution (or that a variables follow a given distribution), but evaluating this time the cdf.

A. *Rayleigh distribution*

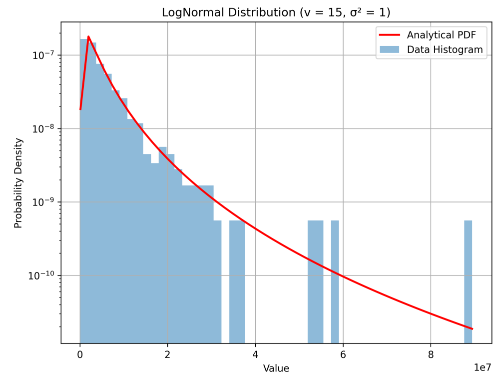
The proposed approach successfully generates the target distribution. The percentage difference between the analytical two momentums and the empirical one are always under the desired threshold. Both *Chi-squared test* and KS-test return a p value greater than the significance level ($\alpha = 0.05$).

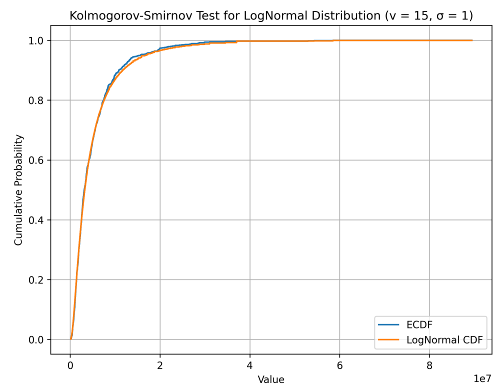




B. *LogNormal distribution*

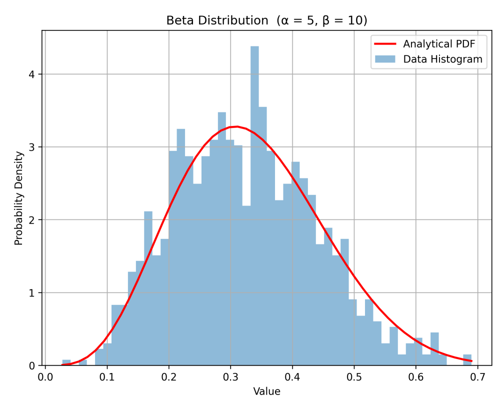
The proposed approach successfully generates the target distribution. The percentage difference between the analytical two momentums and the empirical one are always under the desired threshold. Both *Chi-squared test* and KS-test return a p value greater than the significance level ($\alpha = 0.05$).





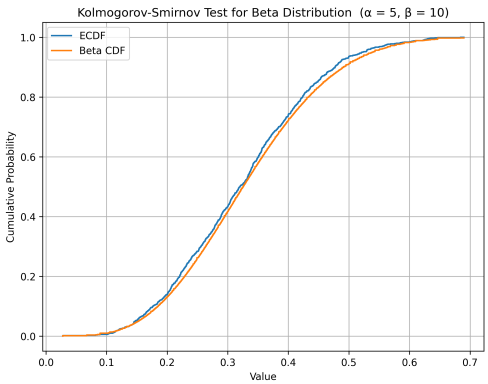
C. *Beta distribution*

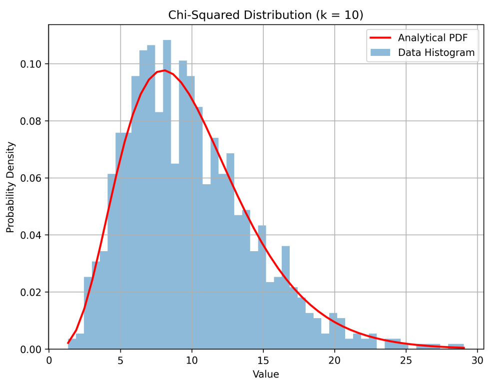
The proposed approach successfully generates the target distribution. The percentage difference between the analytical two momentums and the empirical one are always under the desired threshold. Both *Chi-squared test* and KS-test return a p value greater than the significance level ($\alpha = 0.05$).

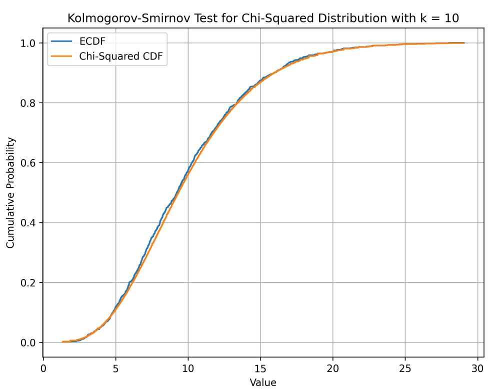


D. *Chi Squared distribution*

The proposed approach successfully generates the target distribution. The percentage difference between the analytical two momentums and the empirical one are always under the desired threshold. Both *Chi-squared test* and KS-test return a p value greater than the significance level ($\alpha = 0.05$).







E. Rice distribution

The proposed approach successfully generates the target distribution. The percentage difference between the analytical two momentums and the empirical one are always under the desired threshold. Both *Chi-squared test* and KS-test return a p value greater than the significance level ($\alpha = 0.05$).

IV. CONCLUSION

The proposed approaches successfully demonstrated that is possible to generate numerous random variable from the Uniform distribution.