Statistics Software Lab Report - 1

Name of the Student: Shatansh Patnaik
Roll No: 20MA20067

IIT Kharagpur
Statistics Software Lab

Generation of Estimation for π using Monte Carlo Simulation

Monte Carlo simulation is a statistical technique that utilizes random sampling for the analysis of complex processes. In this scenario, we will be employing Monte Carlo Simulation to estimate the value of π. The fundamental idea involves generating a set of random points within or on the square formed by the coordinates $(-1,-1)$, $(-1,1)$, $(1,1)$, and $(1,-1)$. Simultaneously, we draw a circle with a radius of 1 and a center at $(0,0)$.

Estimation of the Value of π in R

```r
# Estimation of the value of pi:
BIG_NUM <- 1000
generate_rand_value <- function (num, lo, hi) {
  set.seed(67)
  U <- runif(num, lo, hi)
  return(U)
}

X <- generate_rand_value(BIG_NUM, -1, 1)
Y <- generate_rand_value(BIG_NUM, -1, 1)

count <- 0

for (x in X) {
  for (y in Y) {
    if (x^2 + y^2 <= 1) {
      count <- count + 1
    }
  }
}

pi_val <- 4 * (count / (BIG_NUM * BIG_NUM))

# Approximately the value P(X^2 + Y^2 <= 1) = pi/4,
# so we need to multiply the value by 4 to estimate the value of pi
sprintf("The estimated value of pi is: %f", pi_val)
```

General Algorithm for Solving Random Number Generation Problems under various Probability Distributions

We will be devising a general algorithm for solving all kinds of random generation problems and their successive implementation in R.

Algorithm 1 Chi-Square Test of Goodness of Fit

Input: Probability Density Function $f(x)$

Output: Chi-square statistic, observed values, and expected values

Integrate $f(x)$ to obtain Cumulative Distribution Function (CDF)

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Find the Inverse of the Cumulative Distribution Function (CDF)

$$F^{-1}(u) \quad \text{(where } u \text{ is a uniformly generated random sample)}$$

Update Intervals for Equiprobable Values

Divide the range of $F^{-1}(u)$ into intervals such that values in each interval are equiprobable

Get Observed and Expected Values

Count the observed values in each interval and calculate the expected values based on $f(x)$

Chi-Square Test Statistic

$$\chi^2 = \sum_{\text{intervals}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Output: χ^2, observed values, expected values

Exponential Distribution:

The Exponential distribution is a continuous probability distribution that models the time until an event occurs in a process with a constant rate. It is often used to describe the waiting time between independent events that happen at a constant average rate.

Probability Density Function (PDF)

The probability density function (PDF) of the Exponential distribution is given by:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

where λ is the rate parameter.

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of the Exponential distribution is given by:

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Applying the Algorithm specified earlier to calculate the inverse of CDF, we write the following code in R:

```
sigma <- 67
theta <- 0

# Now we shall calculate the value of the given distribution
get_rand_exp_dist <- function (num){
  x <- runif(num)
  h <- theta + (-1/sigma)*log(1-x)
  return(h)
}

# Now we visualise the exponential distribution:
X <- get_rand_exp_dist(1000)
hist(X, main = "Histogram of Exponential Distribution", xlab = "Value", ylab = "Frequency", col = "lightblue", border = "black")

# First we generate the intervals and k = 8 (let's say)
lambda <- 1000/sum(X)
k<-152
intervals <- numeric(0)
prob<-1/k
E <- rep(1000*prob, k)

for(i in 0:(k-1)){
  intervals<-append(intervals, (-1/lambda)*log(1-i*prob))
}

```

```r
# Now we shall proceed to obtain the observed frequencies
get_observed_freq <- function (X, ints) {
  freqs <-numeric(0)
  for(i in 2 : length(ints)){
    freqs<-append(freqs, sum(X >= ints[i-1] & X< ints[i]))
  }
  freqs<-append(freqs, length(X)-sum(freqs))
  return(freqs)
}

O <- get_observed_freq(X, intervals)
W <- sum(((O-E)^2)/E)
criticial_value <- qchisq(0.95, k-1)
if(W > criticial_value){
  print("The given distribution doesnt follow Exponential Distribution")
} else {
  print("The given distribution follows Exponential Distribution")
}
```

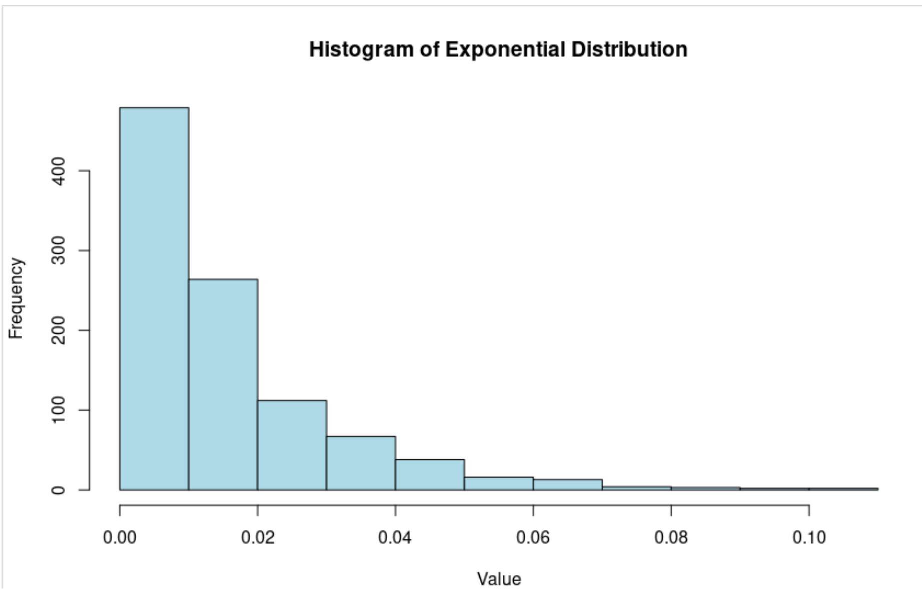

Figure 1: Exponential Distribution generated from a collection of uniform random variables

Standard Normal Probability Distribution (Z)

The Standard Normal distribution, often denoted as Z, is a special case of the Normal distribution with a mean (μ) of 0 and a standard deviation (σ) of 1. It is a central and widely used distribution in statistics.

Probability Density Function (PDF)

The probability density function (PDF) of the Standard Normal distribution is given by the standard normal formula:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

where z is a standard normal variable.

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of the Standard Normal distribution is denoted by $\Phi(z)$ and is given by the integral of the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

The CDF represents the probability that a standard normal random variable is less than or equal to a given value z.

The Following is the code in R for the generation of Random Variables following Standard Normal Distribution:

```r
generate_std_normie <- function(num) {
  x1 <- runif(num, 0, 1)
  x2 <- runif(num, 0, 1)

  z1 <- sqrt(-2 * log(x1)) * cos(2 * pi * x2)
  z2 <- sqrt(-2 * log(x1)) * sin(2 * pi * x2)

  return(cbind(z1, z2))
}

generate_normal <- function(num, mean, stdev) {
  z <- generate_std_normie(num)
  x <- mean + stdev*z
  return(x)
}

mean <- 0
stdev <- 1
num<-1000

sapling <- generate_normal(1000, mean=0, stdev=1)

# Plot histogram to visualize the distribution
hist(sapling, main = "Histogram of Normal Distribution", xlab = "Value", col
    = "green", border = "black")
```

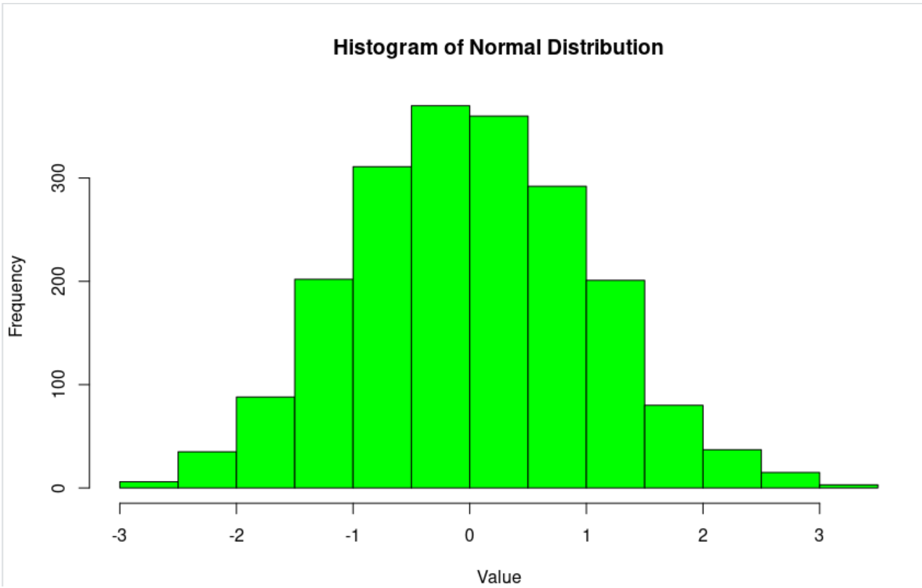

Figure 2: Standard Normal Distribution generated from a collection of uniform random variables

Double Exponential Distribution

The Double Exponential distribution, also known as the Laplace distribution, is a continuous probability distribution that is symmetric and has heavy tails. It is often used in statistics and signal processing.

Probability Density Function (PDF)

The probability density function (PDF) of the Double Exponential distribution is given by:

$$f(x; \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

where μ is the location parameter and b is the scale parameter.

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of the Double Exponential distribution is given by:

$$F(x; \mu, b) = \begin{cases} \frac{1}{2}\exp\left(\frac{x-\mu}{b}\right) & \text{if } x \leq \mu \\ 1 - \frac{1}{2}\exp\left(-\frac{x-\mu}{b}\right) & \text{if } x > \mu \end{cases}$$

```
mu <- 0
beta <- 0.67
get_dde_dist <- function(num) {
  x <- runif(num, 0, 1)
  h <- mu - beta * sign(x - 0.5) * log(1 - 2 * abs(x - 0.5))
```

```
  return(h)
}

X <- get_dde_dist(1000)
hist(X, main = "Histogram of Double Exponential Distribution", xlab = "Value
    ", ylab = "Frequency", col = "red", border = "black")

k<-157
intervals <- numeric(0)
prob<-1/k
E <- rep(1000*prob, k)

for(i in 0:(k-1)){
  intervals<-append(intervals, - beta * sign(prob*i - 0.5) * log(1 - 2 * abs
      (prob*i - 0.5)))
}

O <- get_observed_freq(X, intervals)
W <- sum(((O-E)^2)/E)
criticial_value <- qchisq(0.95, k-1)

if(W > criticial_value){
  print("The given distribution doesnt follow Double Exponential
      Distribution")
} else {
  print("The given distribution follows Double Exponential Distribution")
}
```

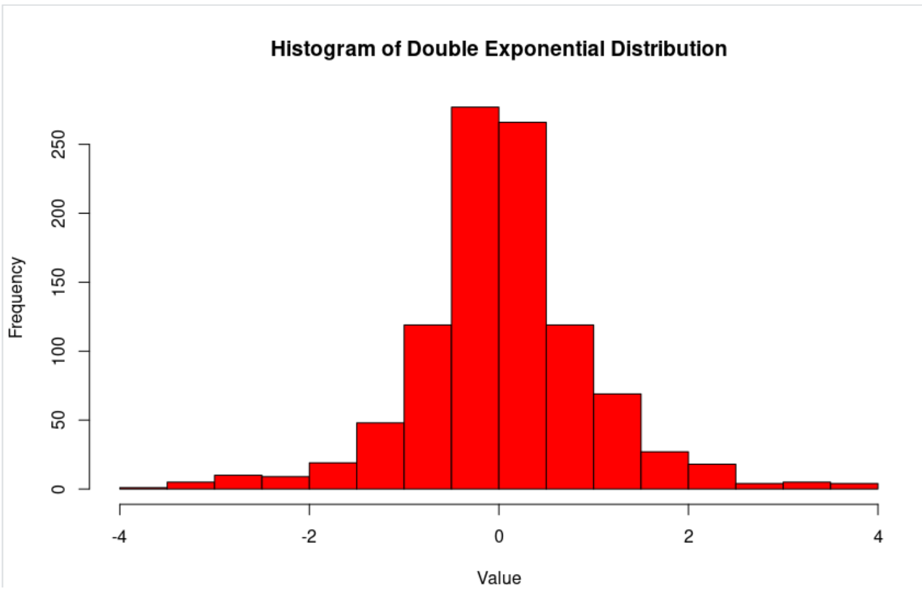

Figure 3: Double Exponential Distribution generated from a collection of uniform random variables

Cauchy Distribution

The Cauchy distribution is a continuous probability distribution that has heavy tails and does not have finite moments. It is often used in physics and statistics.

Probability Density Function (PDF)

The probability density function (PDF) of the Cauchy distribution is given by:

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

where x_0 is the location parameter and γ is the scale parameter.

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of the Cauchy distribution is given by:

$$F(x; x_0, \gamma) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right)$$

```r
sig <- 0.67

get_cauchy_dist <- function(num) {
  x <- runif(num)
  h <- sig*tan(pi*(x-0.5))
  return(h)
}

X <- get_cauchy_dist(1000)
hist(X, main = "Histogram of Cauchy Distribution", xlab = "Value", ylab = "
    Frequency", col = "brown", border = "black")

k<-164
intervals <- numeric(0)
prob<-1/k
E <- rep(1000*prob, k)

for(i in 0:(k-1)){
  intervals<-append(intervals, sig*tan(pi*((i*prob)-0.5)))
}

O <- get_observed_freq(X, intervals)
W <- sum(((O-E)^2)/E)
criticial_value <- qchisq(0.95, k-1)

if(W > criticial_value){
  print("The given distribution doesnt follow Cauchy Distribution")
} else {
  print("The given distribution follows Cauchy Distribution")
}
```

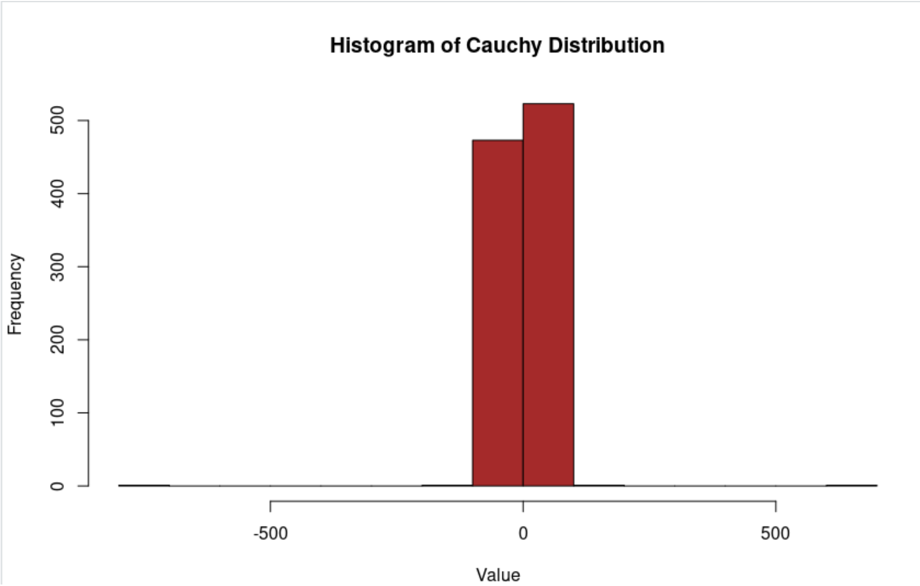

Figure 4: Cauchy Distribution generated from a collection of uniform random variables

Right Trapezoidal Distribution

The Trapezoidal distribution is a piecewise continuous probability distribution defined on a finite interval. It is often used when there is prior knowledge about the distribution shape, and a simple representation is needed.

Probability Density Function (PDF)

The probability density function (PDF) of the Trapezoidal distribution depends on the specific parameters defining the trapezoid. For a trapezoid defined on the interval $[a, b]$ with lower base c, upper base d, and height h, the PDF may be expressed as a piecewise function.

$$f(x) = \begin{cases} \frac{2}{(b-a)(c+d)}(x-a) & \text{if } a \leq x < c \\ \frac{2}{(b-a)} & \text{if } c \leq x \leq d \\ \frac{2}{(b-a)(c+d)}(b-x) & \text{if } d < x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) can be obtained by integrating the PDF over the specified intervals. The specific form depends on the parameters of the trapezoidal distribution.

```
generate_rtz_dist <- function (num, lo, loin, lomax, hi) {
  x <- runif(num)
  h <- numeric(0)
  val <- (lomax - lo) / (hi - lo)

```

```
  for (i in 1:num) {
    if (x[i]<val) {
      h <- append(h, lo + sqrt(x[i] * (loin - lo) * (lomax - lo)))
    } else {
      h <- append(h, lomax + (x[i] - (lomax - lo) / (hi- lo)) * (hi - lomax)
        )
    }
  }
  return(h)
}

lo = 1
loin=1
lomax=1
hi=3

X <- generate_rtz_dist(1000, lo,loin,lomax,hi)
hist(X, main = "Histogram of Right Trapezoidal Distribution", xlab = "Value"
    , col = "red", border = "black")

k<-157
intervals <- numeric(0)
prob<-1/k
E <- rep(1000*prob, k)
val <- (lomax - lo) / (hi - lo)

for(i in 0:(k-1)){
  value <- i*prob
  if (value<val) {
    intervals <- append(intervals, lo + sqrt(value * (loin - lo) * (lomax -
      lo)))
  } else {
    intervals <- append(intervals, lomax + (value - (lomax - lo) / (hi- lo))
        * (hi - lomax))
  }
}

O <- get_observed_freq(X, intervals)
W <- sum(((O-E)^2)/E)
criticial_value <- qchisq(0.95, k-1)

if(W > criticial_value){
  print("The given distribution doesnt follow Trapezoidal Distribution")
} else {
  print("The given distribution follows Trapezoidal Distribution")
}
```

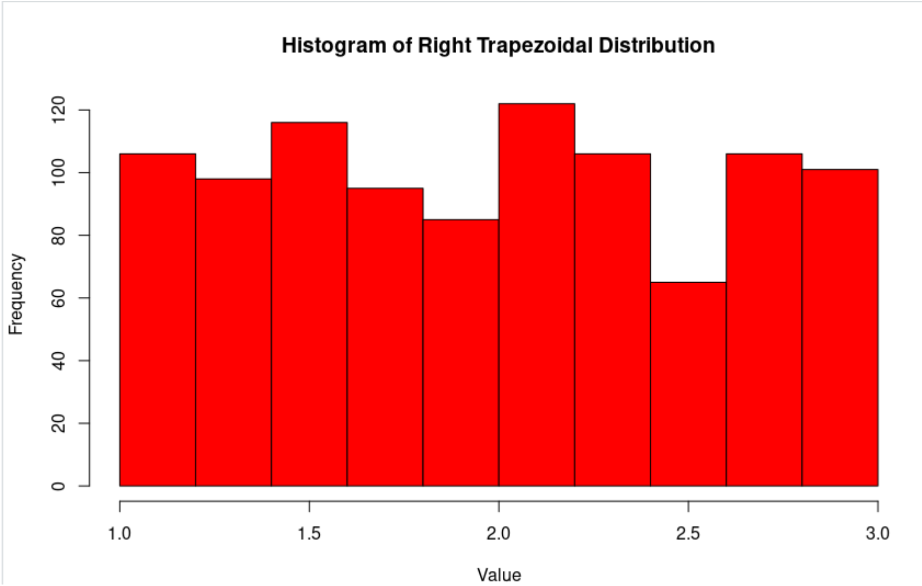

Figure 5: Right Trapezoidal Distribution generated from a collection of uniform random variables

Gamma Distribution

The Gamma distribution is a continuous probability distribution that generalizes the exponential distribution. It is commonly used to model waiting times or durations until a series of independent events occur.

Probability Density Function (PDF)

The probability density function (PDF) of the Gamma distribution is given by:

$$f(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

where k is the shape parameter, θ is the scale parameter, and $\Gamma(k)$ is the gamma function.

Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of the Gamma distribution is not given by a simple closed-form expression, but it can be expressed in terms of the incomplete gamma function.

```
generate_gamma_dist <- function(num, s, r) {
  x <- runif(num)

  generate_gamma_inv <- function(x, s, r) {
    gamma_inc <- function(s, a) {
      integrate(function(q) q^(a-1) * exp(-q), lower = 0, upper = s)$value
    }

```

```
    uniroot(function(x) gamma_inc(x, s) / gamma_inc(r, s) - x, interval = c(0, 100))$root
  }

  h <- sapply(x, function(x_i) generate_gamma_inv (x_i, s, r))
  return(h)
}

s <- 2
r <- 1
num <- 1000

X <- generate_gamma_dist(n, s, r)

hist(X, main = "Histogram of Gamma Distribution", xlab = "Value", col = "blue", border = "black")
```

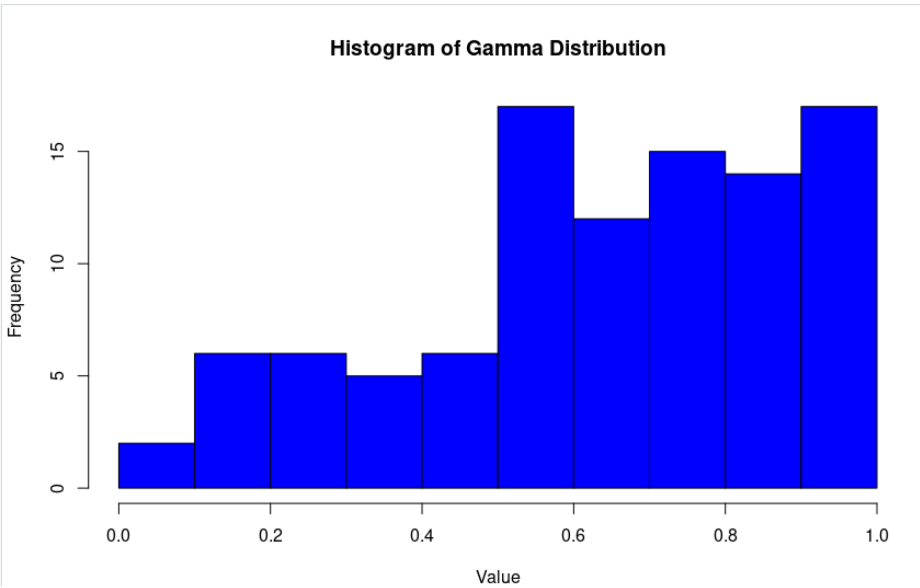

Figure 6: Gamma Distribution generated from a collection of uniform random variables