STA 3431 Assignment #1

Yihan Duan

21/09/2021

Question 1

In this question, we choose another set of parameters for LCG. As discussed in class, by the Hull–Dobell Theorem, we want to make sure:

1. gcd(b, m) = 1
2. every "prime or 4" divisor of m also divides a - 1.

```
# define random function
m = 2^32
a = 4 * 69069 + 1
b = 23606797 * 7
latestval = 12345
nextrand = function() {
    latestval <<- (a * latestval + b)%%m
    return(latestval/m)
}

# record 100000 random variables
rand_vals = c(1e+06)
for (i in 1:1e+06) {
    rand_vals[i] = nextrand()
}
```

As we choose m to be 2^32, we only need b to be odd and a - 1 to be a multiple of 4. To avoid similarities between Un and Un-1, we choose a relatively large a.

Let's see the statistics compared to theoretical values of Uniform(0, 1).

```
# compare statistics with theoretical values
limits= c(100, 1000, 10000, 100000)
for (lmt in limits) {
  cat("For the first ", lmt, "observations:\n")
  cat("Real mean: ", mean(rand_vals[1:lmt]), ", should be: ", 0.5, '\n')
  cat("Real standard deviation: ", sd(rand_vals[1:lmt]), ", should be: ", sqrt(1/12), '\n')
}
```

```
## For the first  100 observations:
## Real mean:  0.4692917 , should be:  0.5
## Real standard deviation:  0.3109622 , should be:  0.2886751
## For the first  1000 observations:
## Real mean:  0.5035736 , should be:  0.5
## Real standard deviation:  0.2943177 , should be:  0.2886751
## For the first  10000 observations:
## Real mean:  0.4992396 , should be:  0.5
```

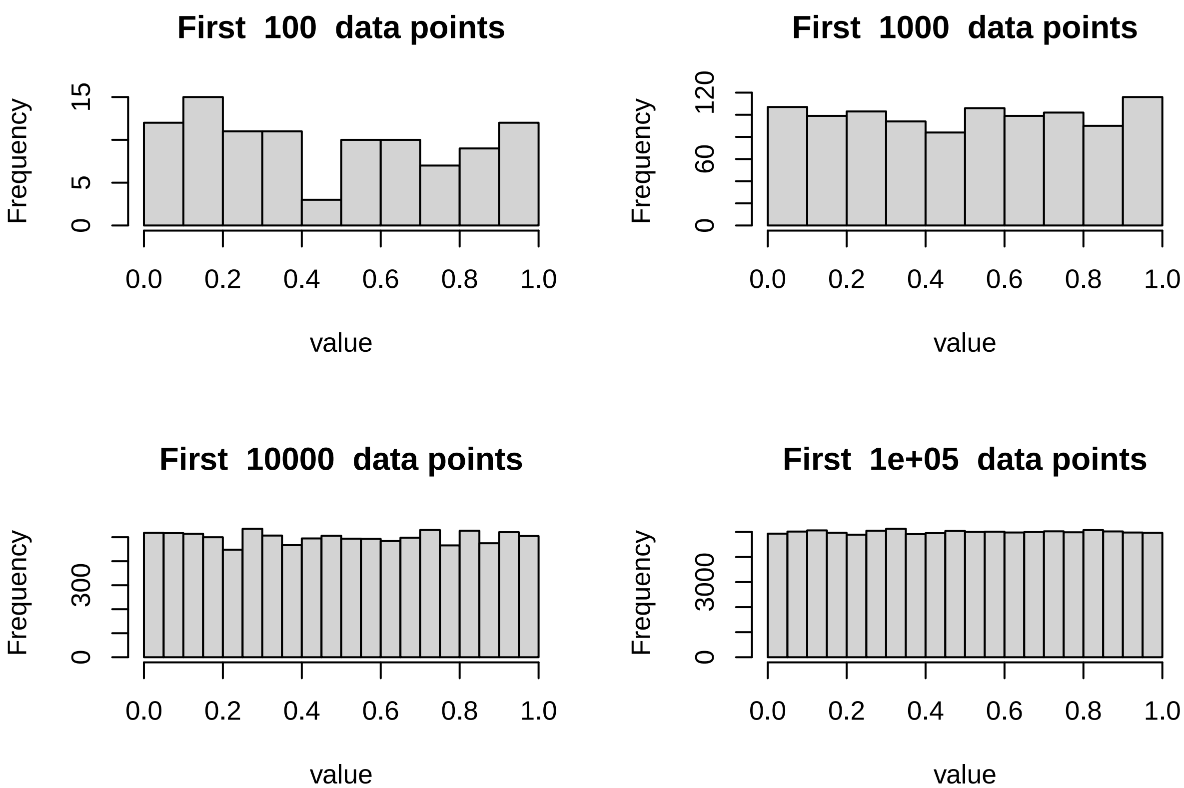
```
## Real standard deviation:  0.2905168 , should be:  0.2886751
## For the first  1e+05 observations:
## Real mean:  0.5003146 , should be:  0.5
## Real standard deviation:  0.2885985 , should be:  0.2886751
```

Randomness

Uniformity

Let's see the distribution of the numbers generated by this algorithm.

```r
# plot frequency of the first N variables.
par(mfrow = c(2, 2))
for (lmt in limits) {
    hist(rand_vals[1:lmt], main = paste("First ", lmt, " data points"), xlab = "value")
}
```



From the histograms alone, we can see that, as the number of observations increases, the values are evenly (uniformly) distributed between 0 and 1. To quantify this, we perform a Chi squared test

Question 2

As discussed in class, we can simulate exponential distribution and normal distribution using uniform distribution.

```r
# simulation size
n = 1000

# function for acquiring n Uniform random numbers
get_randvals <- function(n) {
  rand_vals = c(n)
  for (i in 1:n) {
```

```r
    rand_vals[i] = nextrand()
  }
  return(rand_vals)
}

# function for estimating required expectation
get_exp <- function() {
  # lambda = 3
  U1 = get_randvals(n)
  Y= -log(U1) / 3

  # use Box-Muller transformation
  U2 = get_randvals(n)
  U3 = get_randvals(n)
  Z = sqrt(2 * log(1/U2)) * cos(2 * 3.14159265 * U3)

  # get vars
  X = abs((Y^2) * (Z^5) * sin((Y^3) * (Z^2)))

  # compute and output the mean and standard error
  m = mean(X)
  se = sd(X) / sqrt(n)
  cat("MC:  ", m, " +- ", se, "  (n=", n, ")", "\n", sep='')
  cat("  95% C.I.:  (", m-1.96*se, ",", m+1.96*se, ")\n", sep='')
}

n = 1000000
get_exp()
```

```
## MC:  0.8465307 +- 0.01100011  (n=1e+06)
##   95% C.I.:  (0.8249705,0.8680909)
```

```r
get_exp()
```

```
## MC:  0.8255937 +- 0.01109805  (n=1e+06)
##   95% C.I.:  (0.8038416,0.8473459)
```

```r
get_exp()
```

```
## MC:  0.8367128 +- 0.01068597  (n=1e+06)
##   95% C.I.:  (0.8157683,0.8576573)
```

We can see that the variance is still pretty high. Try increasing the simulation size to see if it gets better.

```r
# Incrase n to reduce standard error
n = 10000000
get_exp()
```

```
## MC:  0.8422648 +- 0.00378406  (n=1e+07)
##   95% C.I.:  (0.834848,0.8496815)
```

```r
get_exp()
```

```
## MC:  0.8319338 +- 0.003634367  (n=1e+07)
##   95% C.I.:  (0.8248105,0.8390572)
```

```r
get_exp()
```

```
## MC:  0.8356876 +- 0.003617307  (n=1e+07)
##   95% C.I.:  (0.8285977,0.8427775)
```

The estimated expectation value is 0.83 with the approximate 95% confidence interval being about (0.82, 0.84). This estimation is not very accurate as the estimated standard error is relatively high (~ 0.0036) even though we are using a large n (10000000). This is because the term $|Y^2Z^5|$ has high variance when $Y \sim$ Exponential(3) and $Z \sim \text{Normal}(0,1)$. In this case, even with large simulation size, the result of "classical" Monte Carlo method still fluctuates a lot between different runs.

Question 3

My student number is 1003118547. Using the last 4 digits, we have:

$$A = 8, B = 5, C = 4, D = 7$$