Plotting a histogram of random numbers generated using a C program

Group - C3
Akilesh Kannan(EE18B122), R Anirudh(EE18B103), Aravint Annamalai(EE18B125)

Using the random numbers that we can generate in the C program - https://courses.iitm.ac.in/pluginfile.php/192607/assignsubmission_file/submission_files/91235/GrpC3_Quiz1_04Sep.c?forcedownload=1 ,this report will deal with the method to plot them in a histogram with an appropriate number of bins, of appropriate size and then checking if they fall into a normal distribution or not, using gnuplot.

As we can see, from the choice of A,C,M values of the recursive relation used in the program, most of the random numbers generated are in the order of 10^9 - 10^{10}, so we can always use a bin width of about 10^8 for plotting the histogram.

Commands to type in terminal to get the random numbers from the output of the C program mentioned above:

//**the following commands are from the beginning, i.e, right from compiling the C program, to getting the output and extracting only the random numbers (excluding mean and standard deviation).**//

```
gcc GrpC3_Quiz1_04Sep.c -o GrpC3_Quiz1_04Sep
./GrpC3_Quiz1_04Sep 25000 1729 | grep -v ' The' > histdata.txt
gnuplot
gnuplot> reset
gnuplot> set key off
gnuplot> set border 50
gnuplot> set boxwidth 50000000 absolute
gnuplot> set style fill solid 1.0 noborder
gnuplot> bin_width = 100000000;
gnuplot> bin_number(x) = floor(x/bin_width)
gnuplot> rounded(x) = bin_width * (bin_number(x) + 0.5)
gnuplot> plot 'histdata.txt' using (rounded($1)):(1) smooth
frequency with boxes
```

The output of the above commands will be:

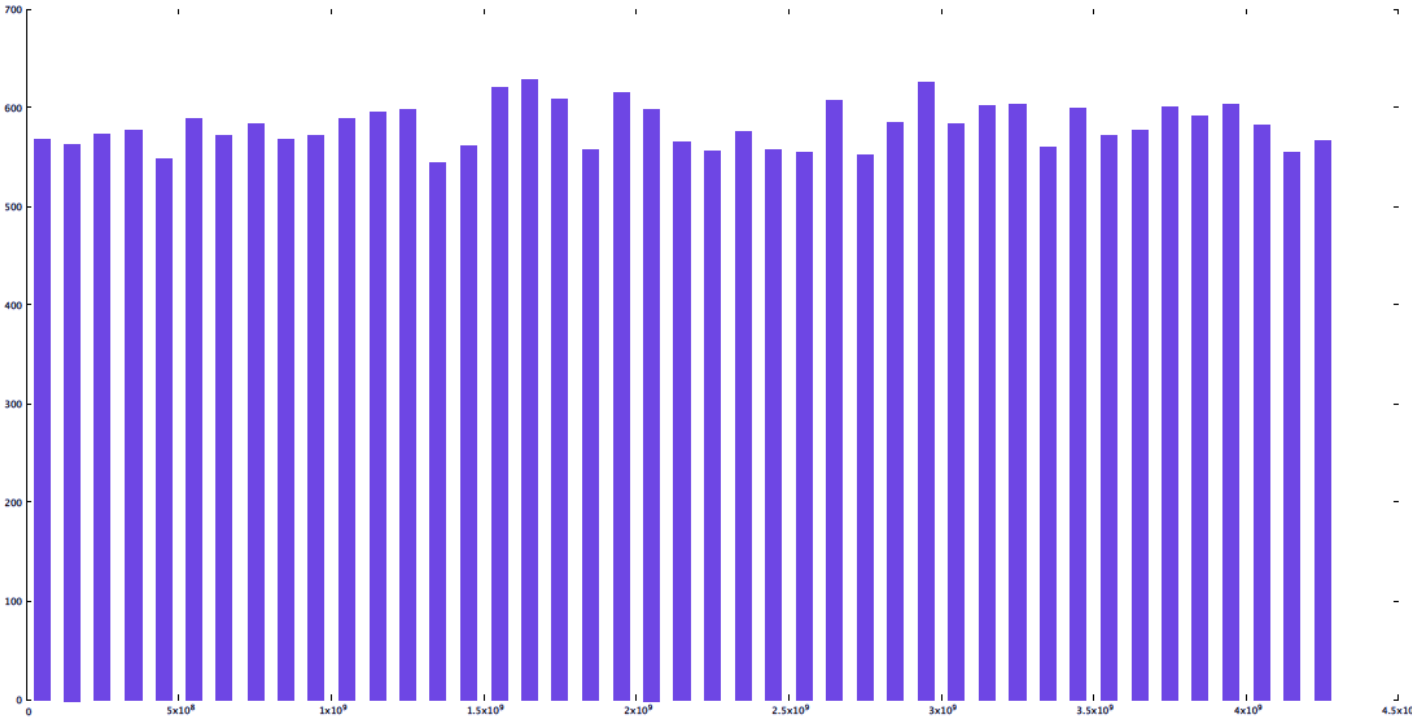

FIG.1 - HISTOGRAM OF THE GENERATED RANDOM NUMBERS

Answer to the 2nd and 3rd subparts of 2nd question:

2 (ii) From the histogram, we can see that it is quite equally distributed over the bins. This shows the truly random nature of the random numbers generated using the recursive relation provided [$I_{j+1} = (A*I_j + C) \pmod M$]. We can also see that all bins are not of the same height. This is because of the fact that we are using a finite number of random numbers. If the number of random numbers used to plot the histogram is very high (of the order of billions or trillions) the height of the different bins of the histogram will almost be equal.

2 (iii) The above histogram does not resemble a normal distribution. This is because we are plotting the random numbers themselves and not their residuals. However, if we plot the value of the algebraic distance (with the sign) from the mean of the random numbers generated, as a histogram, we will get a histogram that looks like a normal distribution.