ITAMARACÁ: A NOVEL SIMPLE WAY TO GENERATE PSEUDO-RANDOM NUMBERS

Daniel Henrique Pereira*

Email: researchdh.pereira@gmail.com

Orcid: https://orcid.org/0000-0003-4750-9659

*Business Administration student at Pontifical Catholic University of Minas Gerais

Abstract: In this paper was presented Itamaracá, a novel simple way to generate pseudo random numbers. In general vision we can say that Itamaracá tends to pass in some statistical tests like frequency, chi square, autocorrelation, run sequence and run test. As an effect to comparison also was taking into account the results of the function RandBetween by Microsoft Excel and true random numbers by Random Org analyzed its distinctive characteristics as well as with the proposal model. In this sense, the goal of this study is contributing to growing the existing Pseudo Random Number Generators (PRNGs) portfolio.

Keywords: Pseudo-random number generator, Itamaracá, Computer Science

INTRODUCTION

According to Knuth (1998), the generation of random numbers has too many practical applications like simulations, sampling, numerical analysis, recreation, computer programming, decision making, and studies on cryptography and aesthetics (computer science), for example.

Although statistical tests do not ensure the generating model is in fact good for practical application, Vieira et al. (2004) raises a series of properties that a “good” random number generator should have in order to be minimally considered acceptable for use. Among these properties we can cite: uniformity, independence, long period, ease of implementation and efficiency, replicability, portability and disjoint subsequences.

Das et al. (2018) states that in most cases, a Pseudo-random Number Generator (PRNG) does not pass all desirable requirements, which is natural due to its deterministic nature. Thus, each PRNG can be used for different types of practical applications, in specific, i.e., some may be useful for the simulation field while other are geared exclusively to the information security field. Anyway, of all these properties mentioned above by Vieira et al. (2004), surely the ones stand out in too many literatures are the uniformity and independence of the numbers generated by the algorithm.

In the present study the PRNG called – Itamaracá - as well as its even more simplified version will be presented. The goal is to present a new simple way to generate random numbers and thus contribute to the increase of the portfolio in this field of study.

1 METHODOLOGY

This study can be considered qualitative and quantitative at the same time. As for the sampling criteria of the generated random number sequences analyzed, convenience sampling was considered.

In this paper, as a way to evaluate the proposed model, a series of tests and statistical tools were considered evaluating, above all, the uniformity and independence criteria.

As for the uniformity criteria, it was used the frequency analysis with graphical tools through histograms; frequency analysis with the chi square test and; Graphical analysis of the distribution of numbers generated through the Scatter Graph.

Also included in the evaluation tests was the Analysis of the mean, standard deviation and repeated numbers of the sequence of random numbers generated in the sample.

As independence criteria, the Autocorrelation function and the Run Test, a binary test considering both odd and even values, as well as values below or above the median were used in this study. Furthermore, the line graph tool was also used to observe the behavior of the numbers generated throughout the series. Another important point, is that also as a way to measure the degree of independence and the level of disorder, Shannon's Entropy was used.

Regarding autocorrelation, the analysis of the values found for the first 10 lags – different of 0 - and their respective average was used.

As a form of analysis, we can say that in the Run Test the first 100 numbers obtained were considered and in the tests containing graphic visualizations (Line Graph, Scatter Plot and Histogram) 1,000 numbers were considered. Furthermore, in the Chi-square Test, Autocorrelation, repeated number analysis and Shannon Entropy, 10,000 number sequences were considered.

In this study, it should also be noted that with regard to data acquisition and manipulation, especially the results generated and data visualization, Microsoft Excel software and the Random Org platform were used.

2 THEORETICAL FRAMEWORK

2.1 Application fields of random numbers

Random number generation according to Rosa (2016) has too many practical applications in our daily lives and in the most different fields of study such as, for example, physics, statistics, computer science, astronomy, astrophysics, medicine. However, the author makes it clear and intelligible that the closest thing to our daily lives is in the fields of data encryption, extremely useful to have a higher level of anonymity of the data of a bank's customers of a patient in a health system, for example.

Moreover, we should also mention the great importance of random numbers for the field of event simulation, from the behavior of demand and its respective profit or loss for a company to; as Rosa (2016) has given us with examples such as the simulation of urban traffic in a large city.

Another social situation in which it is common to encounter random numbers in events directed to the field of leisure and entertainment such as an example the use of random numbers for lottery games, gambling, random selection of a music or video file on smartphones, among many other examples.

2.2 Differentiation between random numbers and pseudo random numbers

According to Rosa (2016) it is natural when we do not delve deeply into the subject, carrying out only common sense, we consider both random numbers (RN) and pseudo random numbers as the same idea. Nevertheless, there are distinctions between these that we must make.

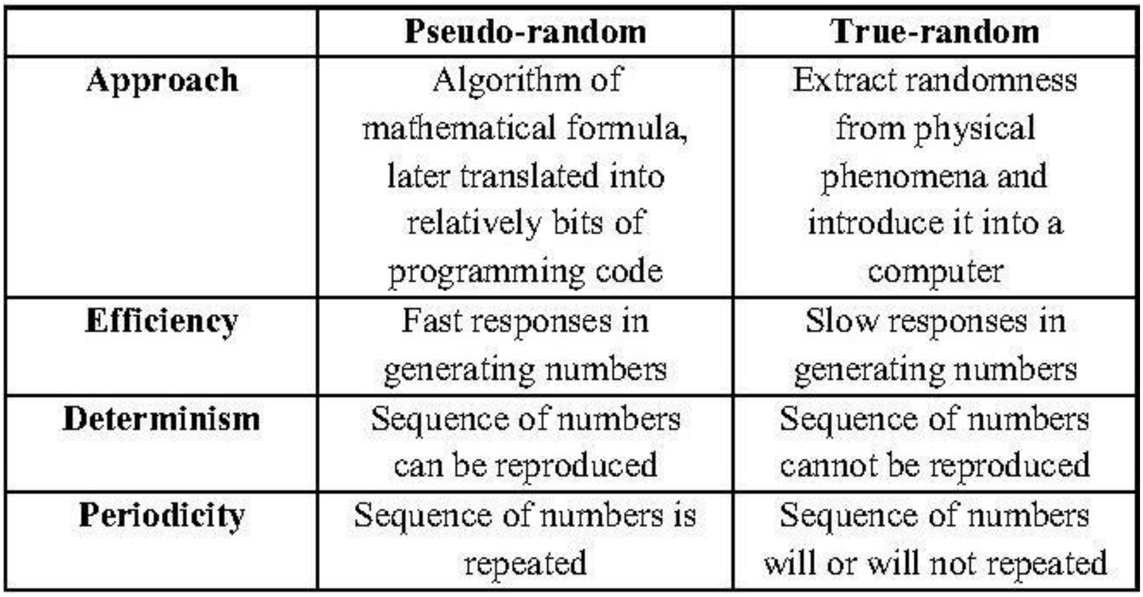
Pseudo random numbers have in their essence, a purely deterministic character, that is, there is a function *f(x)* that given the same initial condition is expected to reach the same results, no matter how many times the experiment is repeated. In physics, this becomes very clear with respect to uniform motion, in which by having data such as time traveled and its respective distance, we can measure the average speed given a unit of measurement during the course of a path.

On the other hand, before we understand what random numbers are, we must first pay attention to what is meant by random. In this sense, we can simplistically say that something is truly random is nothing more than something that occurs without any defined cause, that is, there is no cause-and-effect relationship. Its main characteristic is its unpredictability. In Wikipedia, for example, we are provided with some of its various meanings:

> The expression randomness expresses breakdown of order, purpose, causation, or unpredictability in non-scientific terminology. A random process is the repetitive process whose outcome does not describe a deterministic pattern, but follows a probability distribution (WIKIPEDIA, 2021).

After gaining a little more insight into what randomness consists of, we can say that True Random Numbers (TRN) consist, according to Rosa (2016), of those that possess the aforementioned characteristics which in turn, are usually found in physical phenomena arising from nature, considering everything from the throwing of dice to electrical circuits and radioactive decay, for example.

Moreover, according to Hamid and Abdullah (2015), we should pay attention to the difference between both terms from the point of view of four pillars: approach, efficiency, determinism, and periodicity, as shown in the figure below:



	Pseudo-random	**True-random**
Approach	Algorithm of mathematical formula, later translated into relatively bits of programming code	Extract randomness from physical phenomena and introduce it into a computer
Efficiency	Fast responses in generating numbers	Slow responses in generating numbers
Determinism	Sequence of numbers can be reproduced	Sequence of numbers cannot be reproduced
Periodicity	Sequence of numbers is repeated	Sequence of numbers will or will not repeated

Figure 1. Differences between Pseudo-Random Numbers and True-Random Numbers

by Hamid and Abdullah (2015).

Through Figure 1 by Hamid and Abdullah (2015) we note as for the approach, it will make us clear and intelligible that, in the case of pseudo-random numbers, usually to generate random numbers will be given through some algorithm mathematical formula and therefore, if we consider their applications to the field of computer science, for example, we insert these algorithms so that the computer can understand and reproduce these numbers given a purpose.

On the other hand, when it comes to truly random numbers, before they are inserted into computer language, it is essential the algorithm and/or the generating source of these random numbers come from sources in the natural environment, that is, without any human intervention, such as atmospheric noise, radioactive decay, electrical circuits, among many other examples.

When it comes to efficiency, we see pseudo-random number algorithms tend to be more responsive in generating numbers more quickly and easily. On the other hand, true random number algorithms are not as responsive in generating numbers given their very nature.

Under the pillar of determinism, we also note pseudo-random number algorithms are extremely deterministic, in other words, given the same initial condition we can reproduce the same experiment and generate the same results. However, algorithms of truly random numbers do not have this characteristic, so we cannot obtain the same results even if we start with the same initial conditions, after all, random numbers do not come from mathematical formulations but from some physical phenomenon of nature.

Considering the last pillar, that of periodicity, it becomes clear when it comes to pseudo-random number algorithms, due to their deterministic character, we can soon expect the sequence of numbers can be repeated if given the same initial conditions. As for algorithms of truly random numbers, it is also to be inferred that, most of the time, there is a tendency that there is no periodicity, a repetition of numbers, although in some rare cases due to the random phenomenon, some or other number of a sequence may be repeated by the so-called coincidences.

At this point, it is necessary to conclude and make it understandable the so-called pseudo-random numbers, according to Rosa (2016) are not in fact random. However, as soon as we humans cannot identify and establish patterns in a given sequence of generated numbers, we can consider such numbers to be "random". In this sense, still according to Rosa (2016) not only in common sense but also in too much literature on the subject it is common to trear both terms, pseudo-random and truly random as a synonyms.

2.3 Understanding Pseudo Random Number Generators (PRNG)

Pseudo Random Number Generators (PRNG) are as the name suggests, random number generators given a mathematical function *f(x)*.

It is common, in PRNG models, to come across some basic concepts such as seed, cycle, tail and period.

A "seed" is defined as all values, whether arbitrarily obtained or not that start the process of number generation. As we can see, this is a reference to that which comes from a source and/or that which "is born", "arises". Usually, in the literatures, the seed is expressed mathematically by the letters S.

We can say the "cycle" is the numbers that are generated before repeating sequentially.

The "tail" is the initial part of the numbers that are generated and that normally does not make up the cycle.

Finally, we can say the "period" is the sum of the tail and the cycle. Full sequence of numbers generated.

2.4 Properties of good Pseudo Random Number Generators (PRNG)

Pseudo-random number generators (PRNG) are as their name suggests, generators of random numbers given a function *f(x)*. According to Vieira et al. (2004), there are several aspects that can determine whether a generating algorithm is good or not, such as:

Uniformity: the generating model should be able to generate numbers are very well distributed given the class intervals considered. In this sense, all numbers in a given range must be ½ of probability to came up and therefore, passing statistical tests like Chi Square Test.

Independence: the sequence of numbers considered random y_1, y_2, y_3, yn... must possess the characteristic of independence, that is, for example, not autocorrelated. The next result does not depend on the previous results, at least in principle, thus making it difficult to predict which will be the next numbers generated by this same sequence.

Long period: a pseudo-random number generator must be able to generate a large number of numbers, without repeating themselves. In this respect, preferably without repeating number in cycles.

Ease of implementation and efficiency: in this aspect, it is about the concern in having a generator capable of generating random numbers considering the "computational cost", so, among other aspects, the smaller the amount of algebraic operations the better.

Replicable: pseudo-random number generators must be able to be repeatable, that is, given the same parameters the same results are obtained. This aspect is very relevant for the field of simulations and testing, for example.

Portability: it is about the portability of the generator to handle different types of computers.

Disjoint subsequences: generator must be able to generate new random numbers without therefore going through all the intermediate states.

2.5 Known statistical tools and tests for evaluating pseudo-random number generators

2.5.1 Frequency and Chi-square Test

Stevenson (1981) shows us the chi-square test is an adherence test that is used to evaluate statements made about a distribution in a population.

$$D^2 = \sum_{k=1}^{K} \frac{\left(N_k - m_k\right)^2}{m_k}$$

Figure 2. Chi-square Test formula by Lacerda et al. (2002).

Lacerda et al. (2002) states the chi-square test can be defined as the weighted difference of the observed number of results, or if we prefer, the actual values obtained (N_k), within the interval *k*, and the expected value (m_k) as shown in figure 2 above.

We can say the chi-square test results are good when the *D2* values are as small as possible, especially considering the limits given by the degrees of freedom and desired confidence levels. If, perchance, the values exceed these limits, we can say that perhaps the distribution is not considered well distributed if we consider a uniform distribution, for example.

2.5.2 Linear graphs

Ander-Egg (1971) states that linear graphs are widely used for data visualization practice because of their simplicity of construction on a Cartesian plane. Author also emphasizes although it is not one of the best forms of statistical representation, it can be useful in identifying possible patterns in the behavior of a variable along a generated series.

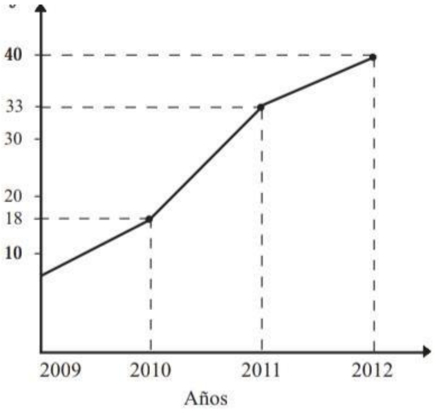


Figure 3. An example of a Linear Graph Image by Bencardino (2012).

2.5.3 Histogram

Histogram according to Levine et al. (2013) is a grouped bar chart containing numerical data, using vertical bars so as to represent the frequencies or percentages in each group, class interval.

Through figure 4 obtained through author Bencardino (2012) we can observe a graphical example of a histogram.

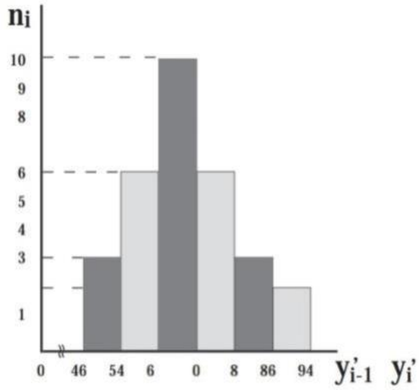


Figure 4. Graphical display of a histogram by Bencardino (2012).

2.5.4 Scatter Plot

A Scatter Plot according to Levine et al. (2013) is a great graphical way to analyze whether there is any relationship between two variables *X* and *Y*, since two-dimensional points are drawn between the variables. Thus, in the field of random number studies, it is an interesting way to analyze the level of uniformity and independency of the data through the points generated between the horizontal and vertical axes of the graph.

About the scatter plot, Bencardino (2012) gives us an example, as shown in Figure 5.

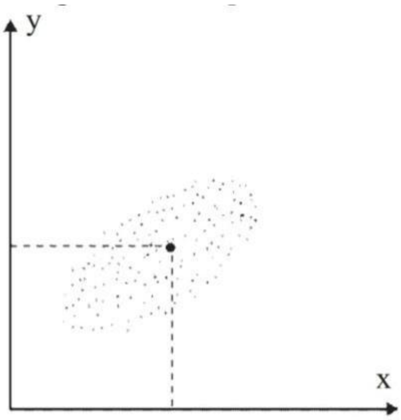


Figure 5. Example of a Scatter Plot by Bencardino (2012).

2.5.5 Average and Standard Deviation

The mean as stated by Levine et al. (2013) is a well-known measure of central tendency with too many practical applications. It serves as a “balance point”, a “neutralizer” given a set of data. We can calculate it by the ratio of the sum of the values by their respective number of values, as you can see in the Figure 6, below obtained through the author Bencardino (2012).

$$\overline{\mathbf{x}} = \frac{\sum \mathbf{x_i}}{\mathbf{n}}$$

Figure 6. Average’s fórmula by Bencardino (2012).

According to Stevenson (1981) standard deviation is a measure of dispersion that is defined as the square root of the deviations from the mean. In addition, the standard deviation is also known as the square root of the variance. In figure 7, we can see how it is calculated.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - [(\sum x_i)^2/n]}{n-1}}$$

Figure 7. Standard Deviation’s Formula by Stevenson (1981).

2.5.6 Run Test

According to Rosa (2016) the Run Test is one of the ways we can evaluate whether or not a given sequence of generated numbers comes from some random process.

Also according to the author previously mentioned, the probability of a number X_n being greater or less than a given value X_i that “separates” the classes follows a Binomial distribution.

Rosa (2016) states that in Run Test it is also necessary to emphasize before calculating the Z-test there are a number of steps we must go through with the data obtained from a generated sequence. First, we must consider a binary character in the separation of the data, that is, numbers “above” or “below” X_i value, where X_i is usually the median value of the complete sequence of generated numbers; in addition, we can create two different categories such as “even numbers” or “odd numbers”, for example.

Suppose that in a sequence of 30 numbers the following distribution of even (E) and odd (O) numbers was observed:

EEEOOEEOEOOOOEOEEEOOOEOEEOOEEEO

The so-called “Number of Runs” of a given sequence of random numbers can be obtained through contiguous groupings as stated by Rosa (2016). In this sense, in this series above, we will find a number of runs equal to 16. Once counted, the number of runs observed will be useful for calculating the expected number of runs, as shown in the figure 8 below.

$$\overline{R} = \frac{2n_1n_2}{n_1 + n_2} + 1$$

Figure 8. Formula to calculate the number of expected runs by Rosa (2016).

After calculating the expected number of runs, we must calculate the standard deviation of the number of sequences, as Rosa (2016) gives us his formulation in the figure below.

$$S_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Figure 9. Standard deviation of the number of sequences by Rosa (2016).

After the previous process, we finally arrive at the Z-test calculation, in which through large samples with n1 > 20 and n2 > 20 as stated by Rosa (2016), we can approximate the statistic with a normal distribution in which we will perform a hypothesis test of, such as:

h_0: sequence of numbers can be considered as random

h_1: sequence of numbers cannot be considered as random.

Z-test can be calculated by the ratio of the difference between the observed number of runs by the expected number of runs divided by the square root of the standard deviation of that same sequence of numbers considered.

$$Z = \frac{R - \overline{R}}{S_R}$$

Figure 10. Z-test by Rosa (2016).

The null hypothesis can be rejected if the z value founded is greater than the critical value as shown below.

$$|Z| > Z_{1-\frac{\alpha}{2}}$$

Figure 11. Rejection of the null hypothesis by Rosa (2016).

2.5.7 Shannon Entropy

Shannon Entropy, a very important concept for the field of Information Theory, was named according to Wu et al. (2012) after studies by Claude Shannon in 1948 through his work "*A Mathematical Theory of Communication*".

This concept can be understood as “a measure of the uncertainty associated with a random variable.Specifically, Shannon entropy quantifies the expected value of the information contained in a message” (Wu et al.,2012, p.2).

Probabilistic Uncertainty Shannon can be defined by the following expression:

$$\mathrm{H}(X) = \mathrm{H}(P_1, \ldots, P_n) = -\sum_{i=1}^{n} P_i \log_2 P_i$$
$$P_i = \Pr(X = x_i)$$

Figure 12. Shannon Entropy Equation by Wu et al. (2012).

2.5.8 Itamaracá

The model proposed in this study - Itamaracá - comes from the Tupi-Guarani language, in which it refers to something like: "*Stone that sings*". In this sense, being, therefore, a reference to something that is "random", "unexpected".

Like every pseudo random number generator, Itamaracá also has some distinctive features in its mathematical algorithm. In this model, we must consider values for 3 seeds, S_0, S_1 and S_2, besides making it clear what the value of N is, that is, the maximum desirable value of "drawn" numbers between the range 1 and N, where where $N \in \mathbb{N}$.

After selected all the 3 seeds, S_0, S_1 and S_2, the calculation process is divided in two main and very simple steps: P_n (n Process) and Final Calculation.

In topics 2.5.8.1 and 2.5.8.2 below it will be shown how the model was originally found in its crude form and that little by little through its “crude” version an even more simplified version was achieved, as shown in topic 2.5.8.3 in view of aspects such as ease of number generation and computational costs.

2.5.8.1 P_n (n Process)

In this stage we need taking into account the absolute values considering the differences between the 3 seeds that must be moving in the sequence.

$$P_n = ABS\ (S_2 – S_1 + S_1 – S_0)$$

Where,

P_n = n Process

ABS = Absolute Value

S_1, S_2 and S_3 = seeds pertenced in the range of *N* selected by a user criterion

After this process, we need to obtain its final result.

2.5.8.2 Final Calculation

In this step, we must multiply the “*x*” result obtained in the first step (in P_n) by the square root value in which its founded value is desirable to be near to 2. That is, so there may be possibilities to generate both even and odd numbers in the sequence and the same time, allow us to generate well distributed numbers within the range of N.

$$FRNS_n = ABS\ \{N - [P_n * SQUARE\ ROOT(1< X_{rn} <4)]\}$$

Where:

$FRNS_n$ = Future values of the sequence of random numbers given a current time period of n.

ABS = Absolute value.

N = Maximum value within a range selected by a user criterion.

P_n = n Process

X_{rn} = Rational number with an arbitrary number of decimal places in the range 1 to 4.

2.5.8.3 Simplified version of Itamaracá

In view of the problematic issue as pointed out in too much literature on computational costs, and that the simpler a PRNG model is the better, an even simpler way to obtain a sequence of random numbers through Simplified Itamaracá will be presented.

The simplified version of Itamaracá, or its final version found and also understood as the closest alternative to computational feasibility, has some features would be unnecessary in its original first version, such as in the n Process (P_n) step the formulation *ABS* ($S_2 – S_1 + S_1 – S_0$) is exactly equal to *ABS* ($S_2 – S_0$). In this sense, we can use *ABS* ($S_2 – S_0$) in the n Process (P_n) step. Therefore, we can arrive at the smallest number of mathematical operations to achieve the same result.

Another relevant point, we can disregard the extraction of the square root of an arbitrarily chosen *x* value between $1 < X_{rn} < 4$ for just a “fixed” value close to 2 from a rational number with decimals also arbitrarily chosen. This obtained value, therefore, will be multiplied by P_n which in turn will be subtracted by N considering absolute values. In this sense, follow the simplified algorithm below:

$$FRNS_n = ABS\ [N – (P_n * X_{rn})]$$

Where:

$FRNS_n$ = Future values of the sequence of random numbers given a current time period of n.

ABS = Absolute value.

N = Maximum value within a range selected by a user criterion.

P_n = n Process

X_{rn} = “Fixed” rational number close to 2 with an arbitrary number of decimal places.

3 RESULTS AND DISCUSSIONS

3.1 Itamaracá Reviews

As a way of exemplifying how the Itamaracá model works, the following introductory values for N and for the seeds S_0, S_1 and S_2 will be considered:

Table 1

Introductory values for the Itamaracá

N = 10,000
S_0 = 4,120
S_1 = 1,300
S_2 = 490

First four numbers generated can be demonstrated below:

1st number:

P_1 = ABS (490 – 1300 + 1300 – 4120) = 3,630

$FRNS_1$ = ABS {10,000 - [3,630 * SQUARE ROOT(3.9)]} = 2,831

2nd number:

P_2 = ABS (2,831 – 490 + 490 – 1,300) = 1,531

$FRNS_2$ = ABS {10,000 – [1,531 * SQUARE ROOT(3.9)]} = 6,976

3rd number:

Here, we can use a "*line break*" between the differences of the seed values. In this sense, P_2 looking like this:

$$P_2 = ABS\ (S_{2n+1} – S_2 + S_1 - S_0)$$

Putting the values into the algorithm,

P_3 = ABS (6,976 – 2,831 + 490 – 1300) = 3,335

$FRNS_3$ = ABS {10,000 - [3,335 * SQUARE ROOT(3.9)]} = 3,415

It should be emphasized that the “line break” is an optional item and may or may not be inserted in the model, at the user’s discretion. At first, statistical results do not tend to differ too much, if not it is just a matter of “hindering” a reversibility of the model and someone arriving at the initial seed values, if tested for something related to data encryption, for example. In this sense, the scientific community is invited to provide more studies on this subject.

4rd number

In the fourth generated number and onward we can return to the initial *Pn*.

P_4 = ABS (3,415 + 6,976 + 6,976 – 2,831) = 583

$FRNS_4$ = ABS {10,000 - [583 * SQUARE ROOT(3.9)]} = 8,848

We can say the first four numbers generated by Itamaracá algortihm were: 2,831 - 6,976 – 3,415 and 8,848.

3.1.1 Frequency and Chi-Square Analysis

As we can see in table 2, the numbers are well distributed, since in each class interval there is an approximate probability of 10% of occurrence, and furthermore, considering the fact that each number within the interval of N also has an approximate chance of 50%, it is therefore perfectly normal that there are some random fluctuations both for more and for less on the average expected probability. In this sense, we can expect in the Itamaracá model, behaviors close to a uniform distribution in the generation of pseudo random numbers.

Table 2

Showing the results of the Chi-Square Test considering 10,000 numbers generated by Itamaracá

n	**Class Interval**		**Freq. (x)**	**Freq. (%)**	**Prob.**	**Prob.* N**	**(A-C)^2/C**
1	0	1,000	1,018	10.18%	1/10	1,000	0.32
2	1,001	2,000	991	9.91%	1/10	1,000	0.08
3	2,001	3,000	994	9.94%	1/10	1,000	0.04
4	3,001	4,000	985	9.85%	1/10	1,000	0.23
5	4,001	5,000	1,000	10.00%	1/10	1,000	0.00
6	5,001	6,000	1,033	10.33%	1/10	1,000	1.08
7	6,001	7,000	1,008	10.08%	1/10	1,000	0.06
8	7,001	8,000	999	9.99%	1/10	1,000	0.00
9	8,001	9,000	999	9.99%	1/10	1,000	0.00
10	9,001	10,000	974	9.74%	1/10	1,000	0.68
Σ			**10,000**	**100.00%**			**2.50**

In this example, in addition to the analysis of the frequency distribution, the algorithm was also tested from the point of view of the Chi Square Test.

Considering a 95% confidence level and dividing our N into 10 class intervals, we obtain a degree of freedom equal to 9. Thus, for the algorithm to be considered valid by the uniformity criterion, the chi square test value cannot exceed the limit of 16.92.

Again, from Table 2. we can see that the Itamaracá result for 10,000 generated numbers was 2.50. Therefore, we can conclude that it also tends to pass the test.

3.1.2 Repeated Numbers

In Itamaracá's algorithm it was identified from the 10,000 numbers generated, 3,694 repeated numbers.

Assuming a pseudo-random number generator has a uniform distribution between 0 and 1 (or, if you prefer, in whole numbers), then it is expected that each number within a given interval has a probability of ½, that is, of 50%.

Thus, it is natural that both PRNGs and TRNGs may to some degree have repeated values.

3.1.3 Run Test (Even/Odd)

As a way of testing the Run Test, we will consider the following first 100 random numbers generated: 2,831 – 6,976 – 3,415 – 8,848 - 6,303 - 4,296 – 1,010 – 453 – 2,412 – 7,232 - 3,386 - 8,076 - 8,333 - 231 - 5,493 - 4,390 - 1,785 - 2,678 - 6,619 - 455 - 5,609 - 8,006 - 4,912 - 8,624 - 8,778 - 2,365 - 2,361 - 2,672 - 9,392 - 3,885 - 7,606 - 6,472 - 4,890 - 4,636- 6,374 - 7,070 - 5,194 - 7,670 - 8,815 - 2,848 - 478 - 6,464 - 2,861 - 5,295 - 7,693 - 457 - 446 - 4,312 - 2,388 - 6,164 - 6,342 - 2,192 - 2,156 - 1,734 - 9,095 - 3,703 - 6,112 - 4,109 - 9,198 - 3,906 - 9,599 - 9,208 - 471 - 8,027 - 7,668 - 4,215 - 2,471 - 264 -2,198 - 9,461 - 8,162 - 1,778 - 5,172 - 4,095 -5,426 - 9,499 - 672 - 613 - 7,548 - 3,579 - 4,141 - 3,271 - 9,391 - 368 - 4,266 - 120 - 9,512 - 359 - 9,528 - 9,967 - 8,975 - 8,907 - 7,906 - 7,889 - 7,990 - 9,834 - 6,158 - 6,383 - 3,184 - 4,126.

In the Run Test a 95% significance level was considered, and with this we obtain its critical value equal to ±1.96, which means values higher than this both positive and negative reject the hypothesis the numbers are considered independent.

Run Test considering alternating between even or odd numbers, the following results were arrived at:

n_1 (Odd) = 41

n_2 (Even) = 59

Number of runs: 50

Expected number of runs = 49.38

Standard Deviation = 23.15

Ztest = 0.128859

3.1.4 Run Test (Median)

Run Test considering the alternation between numbers greater or less than the median, in this example, we obtain a result equal to 5,042. Therefore, we can consider that numbers < 5,042 get a value of "0" and numbers > 5,042 get a value of "1".

Using the same sequence of the first 100 numbers generated by the Run Test (Even/Odd) the following results were obtained:

n_1 (0) = 50

n_2 (1) = 50

Number of runs: 53

Expected number of runs = 51

Standard Deviation = 24.75

Ztest = 0.402015

3.1.5 Autocorrelation

Given a sample of 10,000 generated numbers, we get the following autocorrelation results for up to 10 lags:

Table 3

Autocorrelation observed in Itamaracá through 10,000 generated numbers

k lags	**Autocorrelation (-1:+1)**
Lag 0	1
Lag 1	-0.011056
Lag 2	0.014225
Lag 3	0.010617
Lag 4	-0.006973
Lag 5	-0.020135
Lag 6	-0.008179
Lag 7	0.114837
Lag 8	-0.016052
Lag 9	-0.009193
Lag 10	-0.011365

As we can see in table 3 above, we note the values of the autocorrelation function are in general very close to zero, meaning, according to the literature, a high level of independence. Although in lag 7 we find a slightly higher value than the others, but still when squaring its value, we find a result of 1.32%, that is, 1.32% is the part explained by trying an exercise of generating new numbers.

3.1.6 Some considerations about Itamaracá

Itamaracá - considering both its "crude" version and its even more simplified version in which have the same behavior - in general has proven to be a good random number generator, especially in the criteria that evaluate independence and uniformity. Its applications at values much higher than those demonstrated in this paper have also shown similar statistical results to those obtained through this study. Another point to be highlighted is that it was not observed any rule of choice regarding the value of the seeds, it is enough that they are chosen arbitrarily with values are within the range from 1 to N where $N \in \mathbb{N}$, their maximum value.

As every pseudo-random number generator, Itamaracá also has some identified limitations. As an example, at some point probably after a large amount of generated numbers, the repetition of the same sequence of generated numbers may occur, due to the fact the numbers generated in the sequence are exactly the same and in the same order as the initial seeds (S_2 - S_0) in which seeds are mobile in time. Thus, creating a new cycle of the same numbers generated previously in the period ended.

3.2 Comparing results between the proposed model with the RandBetween Function by Microsoft Excel and TRNG by the Random Org platform

Table 4

Comparing the results between Itamaracá, RandBetween and TRNG by Random Org considering 10,000 numbers generated

	Itamaracá	**RandBetween**	**Random Org**
Chi-Square Test	2.50	8.56	3.65
Repeated numbers/N	3,694	3,653	3,763
Average; Standard Deviation	5,084;2,867	5,005;2,890	4,925;2,905
Run Test (Even/Odd)	0.128859	1.047364	0.004101
Run Test (Median)	0.402015	0.808377	0.603023
Autocorrelation (Average of the first 10 lags different from 0)	0.002827	-0.006046	0.000980
Shannon Entropy	3.45355	3.45355	3.45284

Results obtained considering the same criteria in Itamaracá, we can observe that all the models have passed the frequency distribution test and subsequent analysis of the Chi-square Test, standing out as the lowest value, which is desirable, for Itamaracá.

A second aspect analyzed was with respect to the analysis of repeated numbers in the list given the generation of numbers equal to the value of N, in this case equal to 10,000 numbers generated. At this point, we can observe that all three models have similar results, with a slightly higher value than the others, at first unexpected from the TRNG, but also within normality since all numbers within the range have the same "chances" of appearing given the characteristic of events that follow a uniform distribution.

Regarding the analysis of the mean and standard deviation, we can say that considering a uniform distribution of numbers, the ideal is that they are close to the result of the median, which is equal to 5,042. Thus, we can note all algorithms analyzed are with values very close to the median, highlighting the result obtained by the function RandBetween. Respective standard deviations are also within the expected range, especially when compared to TRNG.

In the Run Test result considering both the alternation between odd and even numbers, as well as the alternation between numbers lower or higher than the median, we see that all models passed the test, since a significance level of 95% was considered and their respective critical value is equal to ±1.96, that is, values higher than this both positive and negative would reject the hypothesis that the numbers are independent.

With regard to the values obtained by Autocorrelation, it is also noted that all three models have values very close to zero. However, standing out as the best result are the values found in the TRNG by Random Org, since due to its nature, we really expect lower values than those found in a PRNG, as is the case of Ita and the RandBetween function by Microsoft Excel.

With respect to Shannon Entropy, we can observe that both Ita and RandBetween - both pseudo-random number generators - have Entropy values very close to the one obtained by TRNG. This is a positive point for both.

Sometimes, despite being within limits, when comparing both Itamaracá – the proposed model – with others such as RandBetween and Random Org TRNG, they may present “better” results than others, which is natural to occur and that changes with each new simulation, with new seeds and new parameters, in the case of a pseudo-random number generator like Itamaracá and RandBetween from Microsoft Excel. As for TRNGs like Random Org, with new data retrieval. In addition, the sample size considered can also impact the results.

Through the results presented above in summary form, we can observe that all of these are in perfect conformity with the statistical tests analyzed.

3.3 Comparison of the graphical visualization results between the proposed model and the RandBetween function by Microsoft Excel and TRNG by the Random Org platform

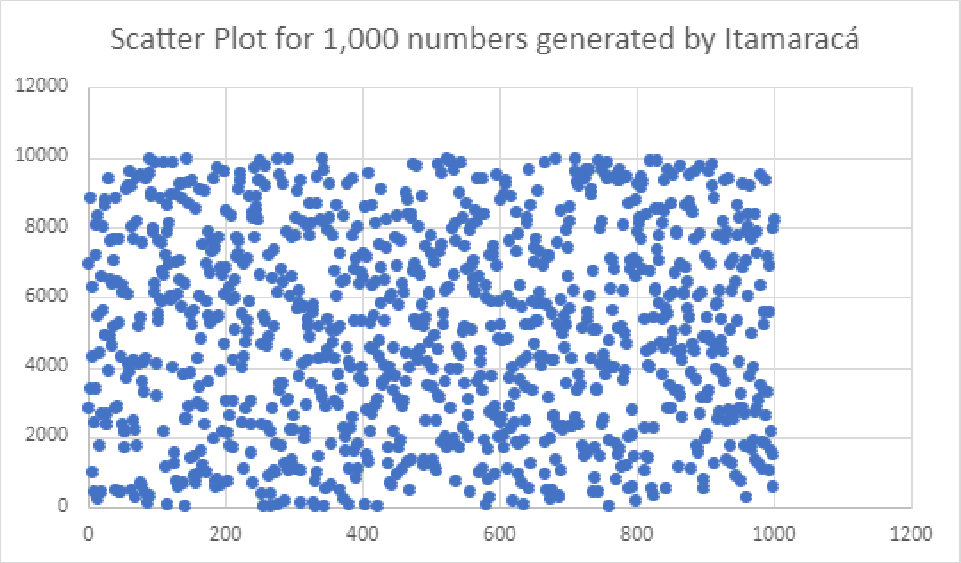


Figure 13. Scatter Plot for Ita.

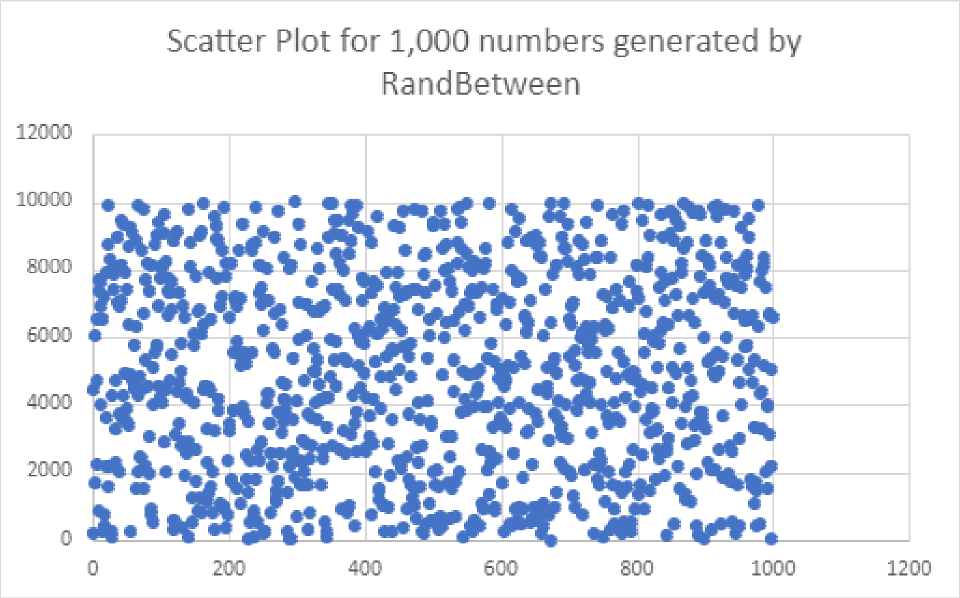


Figure 14. Scatter Plot for RandBetween.

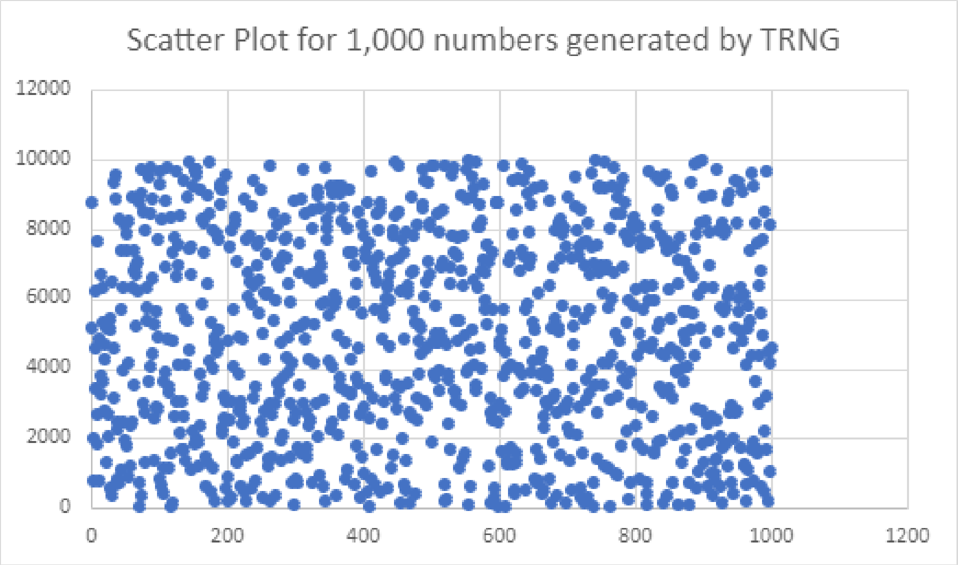


Figure 15. Scatter Plot for TRNG.

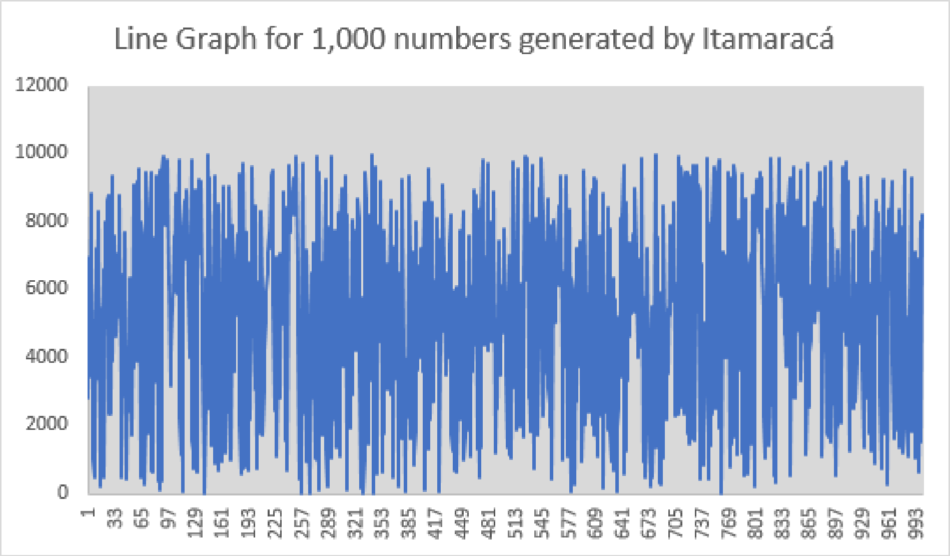


Figure 16. Line Graph for Ita.

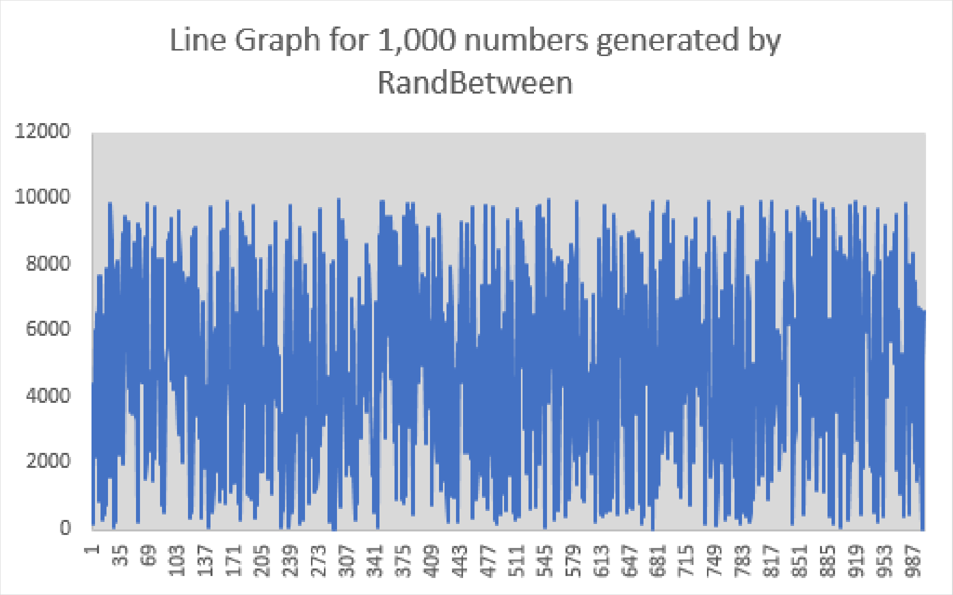


Figure 17. Line Graph for RandBetween.

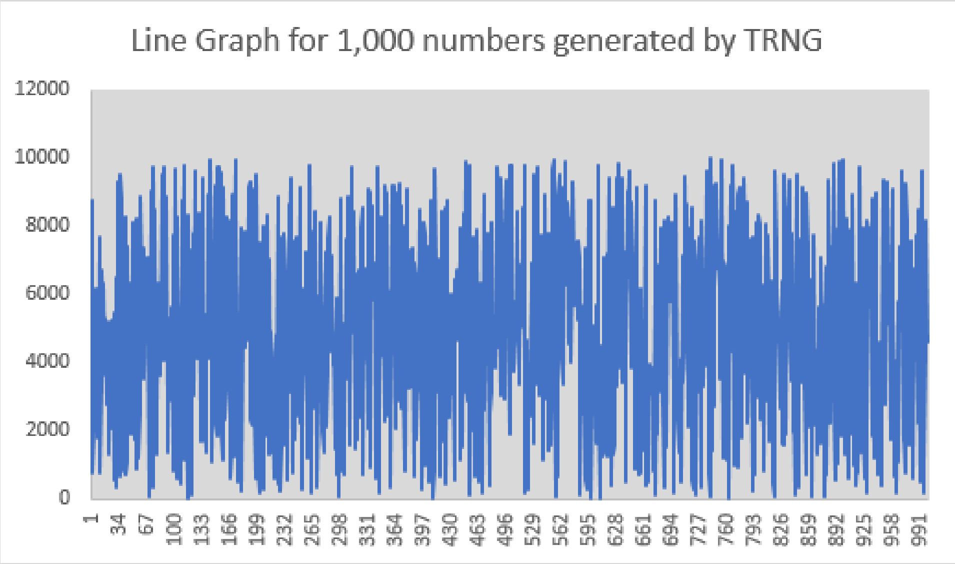


Figure 18. Line Graph for TRNG.

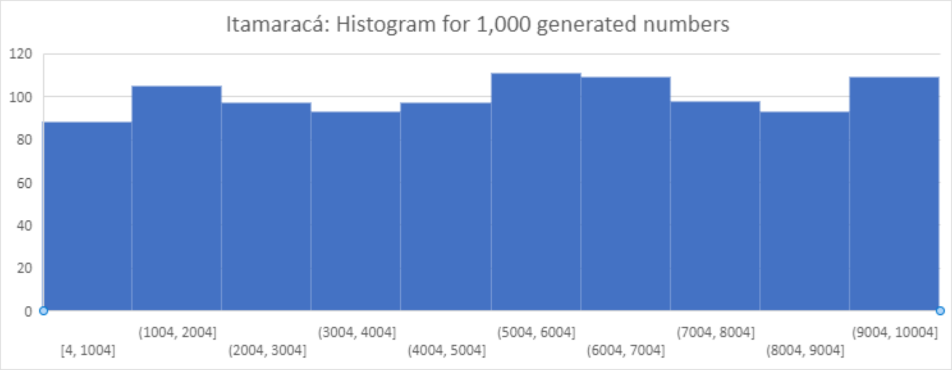


Figure 19. Histogram for Ita.

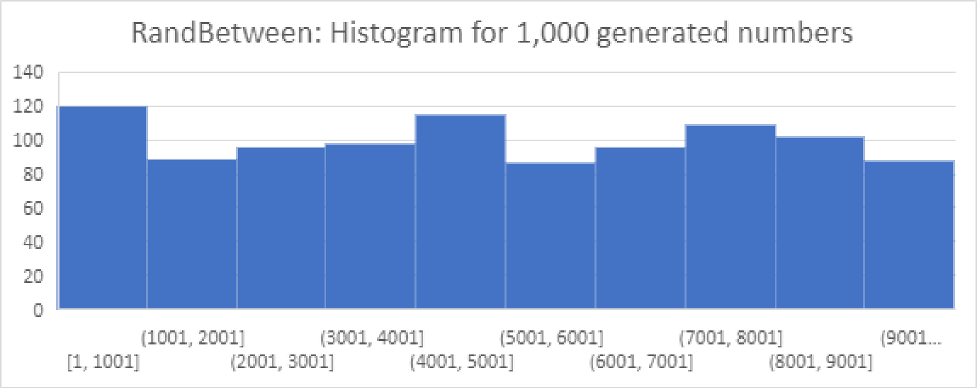


Figure 20. Histogram for RandBetween.

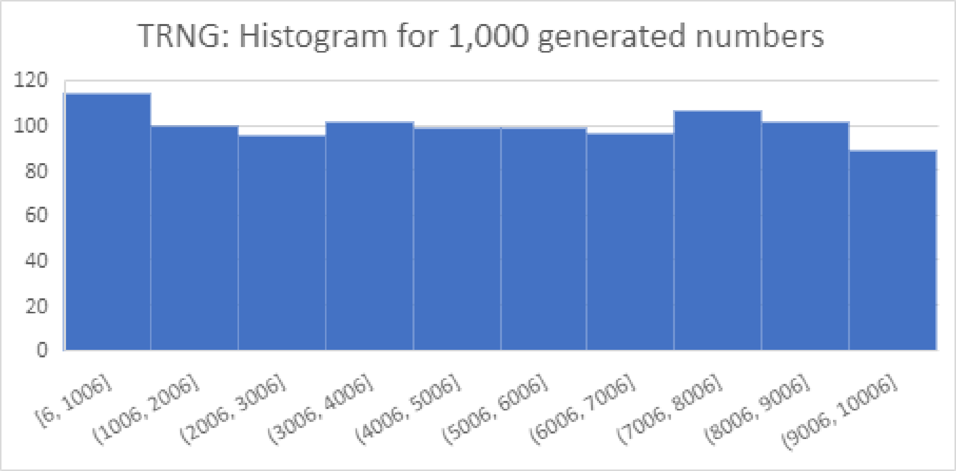


Figure 21. Histogram for TRNG.

We can see from the images in this section all the models compared have features that make the random number generators more reliable.

CONCLUSION

The generation of random numbers is too important for several fields of study and practical applications for the development of mankind. The present study, presented a new and simple proposal of a Pseudo Random Number Generator (PRNG) called "Itamaracá" (Ita in a abbreviated form). Ita model, like all PRNG algorithms, has some limitations, but in general, it showed good results in the statistical tests considered, and thus, as one more model in the portfolio, it is fully available for use and above all, for new studies, especially those applied to a specific objective and real problem.

REFERENCES

Ander-Egg, Ezequiel. *Introducción a las Técnicas de Investigación Social*. [Introduction to Social Research Techniques]. Buenos Aires: Editorial HVMANITAS. 1971.

Bencardino, C.M. *Estadística Básica Aplicada* [Basic Applied Statistics]. Bogotá, D.C.: Ecoe Editions. 2012.

Das, S., Maity, K., Bhattacharjee, K. (2018). *A Search for Good Pseudo-random Number Generators: Survey and Empirical Studies*. [Preprint submitted to Elsevier]. Department of Information Technology, Indian Institute of Engineering Science and Technology.

D. Knuth, The Art of Computer Programming, Vol. 2, SemiNumerical Algorithms. Reading, MA.: Addison-Wesley, 1969.

Herny Ramadhani Mohd Husny Hamid; Norhaiza Ya Abdullah,N "Physical Authentication Using Random Number Generated (RNG) Keypad Based on One Time Pad (OTP) Concept", 2015 Fourth International Conference on Cyber Security, Cyber Warfare, and Digital Forensic (CyberSec)

Kozlowski, L. In *Shannon entropy calculator*. www.shannonentropy.netmark.pl

Lacerda, W.S., Freitas, M.E.A., Pereira, A.R., Jr. Geração de Números Aleatórios [Random Number Generator]. Sinergia, v.3, n.2, p.154-161. 2002. https://www.repositorio.ufop.br/handle/123456789/1643?locale=pt_BR

Levine, D.M., Stephan, D.F., Krehbiel, T.C., Berenson, M.L. *Estatística Teoria e Aplicações usando o Microsoft Excel em Português* [Statistics Theory and Applications using Excel in Portuguese]. Rio de Janeiro: LTC. 2013.

Randomness. (2021, October 27). In *Wikipedia*. https://en.wikipedia.org/wiki/Randomness

Rosa, C.A. (2016). *Números Aleatórios: Geração, Qualidade e Aplicações* [Random Numbers: Generation, Quality and Applications] Universidade Federal do ABC.

Stevenson, W.J. *Estatística Aplicada a Administração* [Statistics Applied to Management]. São Paulo: HARBRA. 1981.

Vieira, C.E.C., Ribeiro, C.C., Souza, R.C. *Geradores de Números Aleatórios* [Random Number Generators]. Rio de Janeiro: PUC Rio. 2004.

W. Yue, Y. Zhou, G. Saveriades, S. Agaian, J. P. Noonan, and P. Natarajan, ''Local Shannon entropy measure with statistical tests for image,'' Inf. Sci., vol. 222, no. 222, pp. 323–342, Feb. 2013.