# INVESTIGATION INTO THE ROLE OF APPROXIMATE ENTROPY TESTING FOR RANDOM NUMBER GENERATORS

AUTHOR KARL MÖLLER 18005742—SUPERVISOR BRYCE ANTONY—MODERATOR BRIAN CUSACK **MATH705 2020**

## INTRODUCTION

OBJECTIVES & PURPOSE

The goal of this project has been to investigate the methods of testing commonly utilized in Random number testing. After an initial phase of research it became apparent that a lot of the testing methods have not seen much of a review since Juan Soto wrote a paper describing the National Institute of Standards and Technology (**NIST**) test battery. Further research indicated that **approximate entropy** testing has been one of the areas that has not been reviewed recently for testing random numbers. Therefor the goal of this project became orientated towards determining whether approximate entropy testing should still be a viable measure for determining randomness in a generated number sequence.

## BACKGROUND

LITERATURE REVIEW

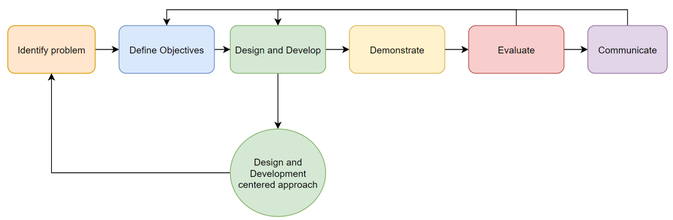
In order to determine the role of approximate entropy testing in the testing of Random Numbers. Research was conducted into the vast list of tests available for methods of testing random numbers. Approximate entropy testing in particular was interesting as it directly addresses regularity. (Soto, 1999) Approximate entropy testing was developed by Steve M. Pincus to address limitations found in traditional methods of regularity testing that relied on vast amounts of data, these large quantities of data would often be skewed or distorted by system noise bringing a dilemma when used with experimental data sets, in that it just was not practical. Approximate entropy (**ApEn**) was developed originally for medical purposes such as measuring regularity in heart rates but over time was expanded out to other fields (Pincus, Gladstone, & Ehrenkranz, 1991). This method of measuring regularity gained popularity due to it's lower computational cost than previous methods, and was also less affected by system noise.

In order to determine randomness other measures were implemented. Two particular tests are noteworthy in relation to ApEn due to the kind of measures they return. Spectral Analysis and Cumulative Sums testing. Spectral analysis identifies periodic features in a sequence, and cumulative sums checks for concentration of 0's or 1's at the beginning of a sequence. These tests are noteworthy as they, like approximate entropy, measure regularity to some degree, therefor they would be viable statistics to prove the role and efficacy of approximate entropy testing.

By checking the start of a binary Random number sequence (RNS) for long strings of 0's or 1's, insight is given into the regularity of a generator over a larger data set. For instance if a sequence regularly presents long strings of 1's or 0's it will regularly fail the test. The same logic applies to spectral analysis. If periodic frequencies are more prevalent than others this implies regularity of those periodic frequencies.

## METHODOLOGY

DESIGN SCIENCE RESEARCH METHODOLOGY




Applying a **Design Science Research Methodology (DSRM)** to this project enabled a simple work flow that led to the production of several artefacts. These artefacts include the various methods used to portray the statistical findings from the experimental data that has been collected. Following a design and development centred approach the collected data was manipulated and presented through several different graphs and tables in order to come to a final form that held in it the conclusions of the research and the answers to the general investigation. Advice was sought and adjustments were made to objectives and the design of the final artefacts. (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007)

## EXPERIMENT

HYPOTHESIS TESTING

RQ: Which RNS generated by PRNG produce random number sequences that pass tests for randomness?

SQ1: Do naturally occurring number sequences pass tests for randomness?

H0: That RN generated by PRNG will pass the NIST guidelines for testing randomness within number sets.

H1: That naturally occurring Random number data sets pass the NIST guidelines for testing number randomness.

H2: That the test strings produce the same results when compared with different NIST random number testing tools.

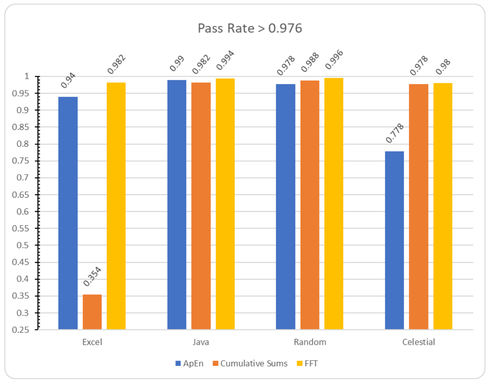
Testing of these hypothesis has given insight pertaining to the research question and sub questions.

Three separate tests were carried out on each binary RNS, five hundred samples of 10,000 bits were extracted for each distinct test and RNS, resulting in 1500 separate p-values that surmise the strength of the H0. Using the NIST standards α is a tolerance level that is equal to 0.01. for any p-value >= α we can determine that a sequence passes the threshold to be considered random. A p-value of 0 indicates complete non randomness as there is no probability that the sequence tested will pass a test for randomness and is therefor not random. A p-value of 1 would therefor give a 100% probability that a sequence is random.

The 3 tests were conducted on 4 distinct data sets. They are Microsoft excels random function, Java.Utils random object, and the Random.org online API that utilizes atmospheric data as a source of entropy, as well as some celestial data provided by Auckland University of Technology's astronomy department.

## RESULTS

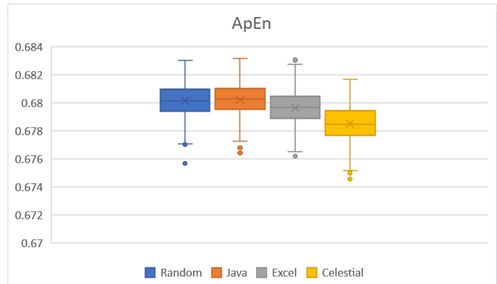
WHAT HAS BEEN ACHIEVED



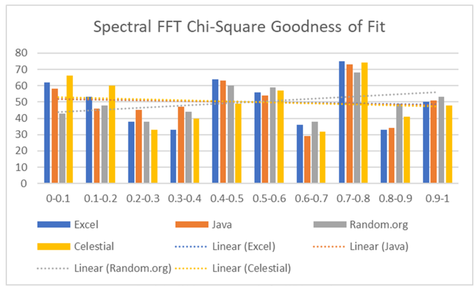

Graph 1—pass rates of generators on the whole for each test

This graph shows the general pass rate as a percentage. When the 500 sequences are put through each test they return a p-value, if the p-value is less than 0.01 they fail the test and subsequently the sequence is defined as non-random. On this graph any generator that does not pass 97.6% of the tests will instantly fail the NIST standard for randomness. There are 3 general failures in this data set. The excel numbers failed the ApEn test and the Cumulative sums test and the celestial data failed the ApEn test.

Now given the results displayed by the pass rate it is important to determine the regularity of the tests. Primarily how well did the ApEn test actually show the regularity of the number sets. Given the pass rate we would expect that the celestial data would have considerable lower values of ApEn than say java. The graph 2 indicates the distribution of the individual results pertaining to ApEn. However when using the other 2 tests the celestial data has a relatively uniform distribution of p-values when evaluating goodness of fit through a Chi-Square test. As shown in graph 3.




Graph 2—Box and Whisker representation of distribution of ApEn values showing a comparative performance




Graph 3—Chi-Square test results for FFT Spectral test

These results indicate that there may be more to see here and further research may be necessary to determine why the ApEn test has reported in such an extreme way in comparison to the spectral test which measures periodicity that should hold a correlative relationship with regularity. It is also worth noting that Graph one indicates a significantly higher fail rate than anything else for excels cumulative sums testing. Supporting a claim that although regularity may be present in the start of a sequence it does not automatically disqualify it from performing to some degree in a test of ApEn. Although the excel data did on the whole fail both the ApEn test and the Cumulative sums test.

## CONCLUSION

FURTHER RESEARCH

To address the RQ we can determine that in general 2/3 of the tested RNGs passed the tests for randomness to the standard that **NIST** seeks. SQ1 is a direct question towards the celestial data collected. And it is determined that of the 3 tests performed the celestial data set passed 2/3 of those tests to the standard expected by NIST. Further research is required to determine the purpose of this investigation that seeks the role of ApEn testing in determining randomness but it could be concluded based on the results of these experiments that the Implicit reasoning that a higher ApEn value results in irregularity is not truly the only measurable feature of Randomness. And other features may be present in a RNS that need to be isolated beyond their irregularity such as their distribution of periodical features that can be represented by waves using a FFT Spectral test. Further research should be done into the particular role of regularity as a feature of randomness to give a more definitive understanding of the role of Approximate entropy testing in determining randomness. As it currently stands the only real conclusion that can be given is the same conclusion Soto came to in 1999 that approximate entropy "detects the non uniform distribution of m-length words." (Soto, 1999) This distribution is where the key role of ApEn could be found when determining randomness and gives reason as to why it has not simply replaced the FFT Spectral Analysis or another test that detects regularity of periodic time series data sets

**References**

Peffers, K. E. N., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45-77. doi:10.2753/MIS0742-1222240302

Pincus, S. M., Gladstone, I. M., & Ehrenkranz, R. A. (1991). A regularity statistic for medical data analysis. *J Clin Monit, 7*(4), 335-345. doi:10.1007/bf01619355

Soto, J. (1999). *Statistical testing of random number generators.* Paper presented at the Proceedings of the 22nd national information systems security conference.

Karl Moller: Linkedin