Distribution of Digits in Integers

Donald McCrae, Garrett Rangel, Elizabeth Striker, and Julia Wright
Advised by Dr. Barbara Margolius - Cleveland State University

Abstract

Probability is a useful tool with many applications. Our project applies probabilistic techniques to the study of the distribution of digits in integers and is based on the journal article 'Distribution of Digits in Integers' by Ralph P. Boas, Jr. Boas' analyzed the distribution of digits in infinitely large integers to determine whether they tend toward normal, meaning roughly that each digit is equally likely to appear in the number regardless of the base. He found that there are many large integers which have two digits with distributions that follow his definition of normal, but it is unlikely that more than two digits will follow the same rule. Our research expands on Boas' work by using probability tests to analyze the distribution of digits in very large integers of different bases to determine whether they follow his observations. We also analyzed the distribution of digits in e, π, and √2 to determine whether they appear normal by Boas' definition.

Background

It is well known that probability and randomness is a significant part of today's society, from making a simple decision using a random number generator online to randomized draft picks for a fantasy sports team. In this project, we set out to determine how "random" these outcomes actually are based on integers. Basic math teaches us that if we have ten numbers and randomly draw one, the probability of any one of those numbers being chosen is 10%. Do irrational numbers such as e, π, and √2 have digits with normal distributions? What about numbers larger than base ten? These are questions we set out to answer.

OBJECTIVES

- Examine whether or not digits in randomly generated numbers in base 10 are normally distributed.
- Use a program we created to randomly generate numbers to run a chi square test for goodness of fit on the digits.

METHODS

- We randomly generated numbers of 1 million digits using a random number generator and a program.
- Next, we ran a chi square test on each number for the digits 0, 1, and 2 to determine if each digit is normally distributed.
- Lastly, we examined the p-values and determined that the results we found differed from the results from the reference article.

P-values

Random Number 1: .829 (no significance)
Random Number 2: .462 (no significance)
Random Number 3: .651 (no significance)
Random Number 4: .762 (no significance)
Random Number 5: .134 (no significance)
Random Number 6: .665 (no significance)
Random Number 7: .023 (significant)
Random Number 8: .217 (no significance)
Random Number 9: .845 (no significance)
Random Number 10: .633 (no significance)

Program Number 1: .167 (no significance)
Program Number 2: .795 (no significance)
Program Number 3: .713 (no significance)
Program Number 4: .614 (no significance)
Program Number 5: .67 (no significance)
Program Number 6: .204 (no significance)
Program Number 7: .813 (no significance)
Program Number 8: .518 (no significance)
Program Number 9: .964 (no significance)
Program Number 10: .133 (no significance)

RESULTS

- After running the chi square test on all the randomly generated numbers, we found that all of the numbers except for one had statistically insignificant results.
- Each digit in each number that was generated was normally distributed, which disagrees with Boas's claim that no more than two digits in a number will tend towards a normal distribution.

Program Trials

1,000,000 digit numbers, 1000 Trials

Number :	Count	Percentage
Zero (0):	100002192	10.0002192
One (1):	99995481	9.9995481
Two (2):	100009455	10.0009455
Three (3):	100013409	10.0013409
Four (4):	99989585	9.9989585
Five (5):	99998440	9.9998440
Six (6):	99993419	9.9993419
Seven (7):	99991297	9.9991297
Eight (8):	100005131	10.0005131
Nine (9) :	100001591	10.0001591

Linear Congruential Generator

$$X_{i+1}=(a*X_i+c) \bmod(m)$$

CONCLUSIONS

The p-values calculated from the distributions were consistently insignificant, except for one trial. This demonstrates that the digits of the large random integers were evenly distributed and normal by Boas' definition, with each digit of base 10 occurring about 1/10 of the time. Therefore, we proved Boas's observation incorrect. Our results suggest that as integers become very large, they have each of their digits evenly distributed.

FUTURE WORK

The trials with our custom java program and various online random number generators showed us that the generators result in an even distribution. For future work, we could observe distributions of different numeric bases with very large random numbers. These studies would all give insight on how truly random a number generator is.

References

- Boas, R. P., Jr. (1977, September). Distribution of Digits in Integers. *Mathematics Magazine*, *50*(4), 198-201.
- Random Digits Generator: https://numbergenerator.org/randomdigits#!numbers=1&length=1000000

Acknowledgments

Dr. Barbara Margolius