# Digit Distribution of Prime Numbers in Binary

Stefanie Wang and Jim Sauerberg
Saint Mary's College of California

## Abstract

The distribution of prime numbers is thought to be random. Hence, if the primes are written in base *b,* we would expect the digits in their base *b* expansions to occur with equal frequency. This seems not to be the case; in fact, for each *b* the smaller digits seem to be more prevalent.

## Introduction

The nature of the distribution of prime numbers remains a mystery to this day. What of their bits, the zeroes and ones, in binary? Would the digit distribution appear to be random, similar to the distribution of primes? Would there be a bias toward a particular binary digit?

This road of inquiry led to the path of the mechanics of the project. Binary representation writes any integer as a sum of powers of two. The binary bits from the right start with $2^0$ and grow sequentially bigger toward the left. Other than two, prime numbers in binary form must begin and end with one, so only the middle terms are of interest to us. For example, seventeen in binary is 10001. Discarding the first and last bits, 17 then becomes 000. Here are some more examples:

| | | | | | |
|---|---|---|---|---|---|
| 11=1011→01<br>13=1101→10 | Start<br>$2^3$=8 | End<br>$2^4$=16 | Primes<br>2 | 0's<br>2 | 1's<br>2 |
| 17=10001→000<br>19=10011→001<br>23=10101→010<br>29=11101→110<br>31=11111→111 | Start<br>$2^4$=16 | End<br>$2^5$=32 | Primes<br>5 | 0's<br>7 | 1's<br>8 |
| 37=100101→0010<br>41=101001→0100<br>43=101011→0101<br>47=101111→0111<br>53=110101→1010<br>59=111011→1101<br>61=111101→1110 | Start<br>$2^5$=32 | End<br>$2^6$=64 | Primes<br>7 | 0's<br>13 | 1's<br>15 |

This brings up the question: are there more ones or zeroes in these binary representations of primes?

## Methodology

The first step in this project was to find the prime numbers between $2^0$ and $2^{n+1}$. Once the primes were found, they were written in binary. Primes were sorted into exponential groups, i.e. between $2^n$ and $2^{n+1}$ for various *n*'s. The total number of zero bits and one bits in each group were computed.

The challenging part of the process was finding the prime numbers since there is no list of primes. The Miller-Rabin Test was used in this process with bases 2, 7, and 61 which weed out all pseudoprimes less than 4.7 billion.

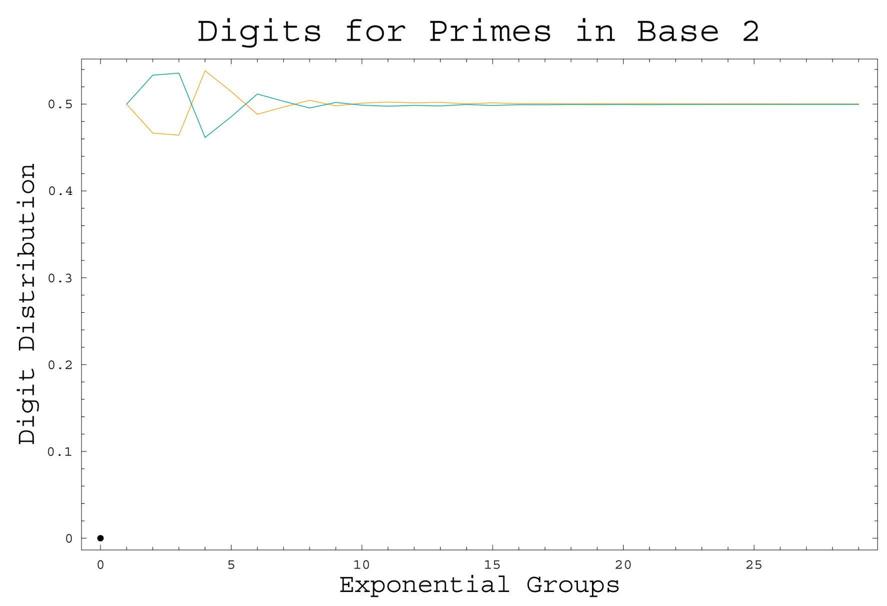



**Figure 1 - Digit Distribution for Primes in Base 2**
The distribution in binary shows a 50/50 distribution for zero and one bits. The green represents the percentage of ones and the orange represents the zeroes. Though the percentages fluctuate initially, they quickly converge.

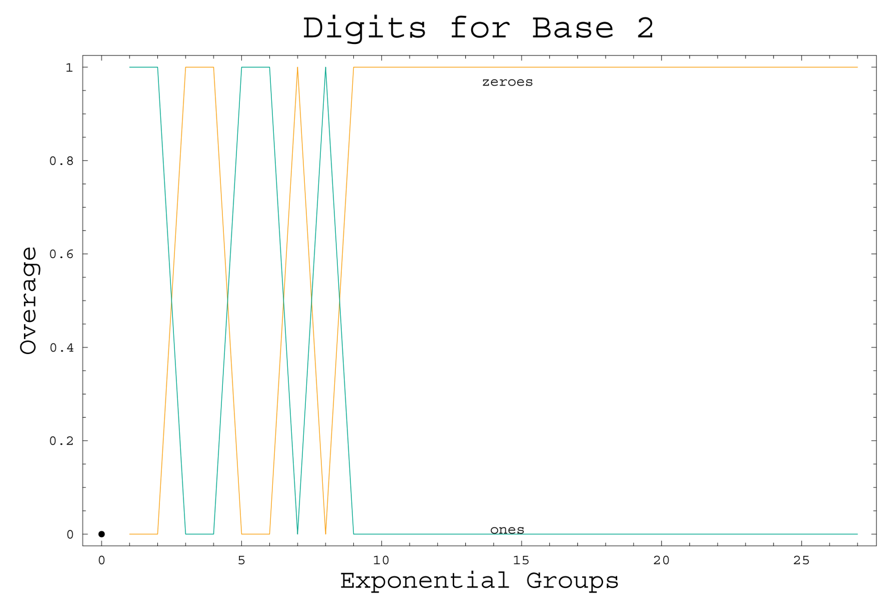
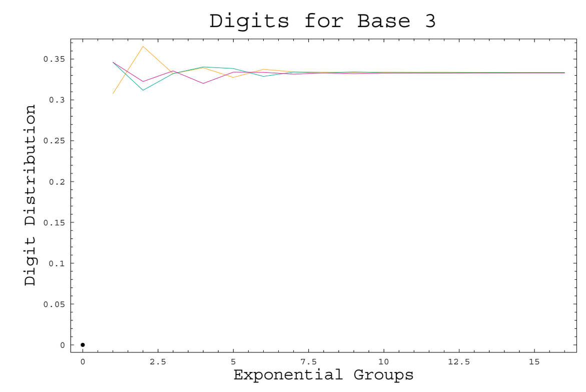


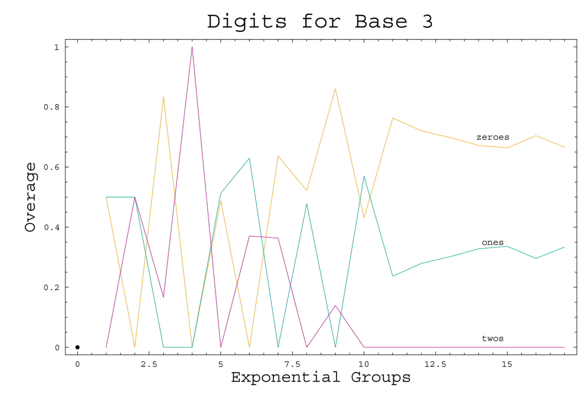

**Figure 2 - Digit Overage for Primes in Base 2**
After $2^{12}$ there seems to always be more zeroes than ones. This graph shows the relative "error" of the digits in each group when compared with the expected uniform distribution.

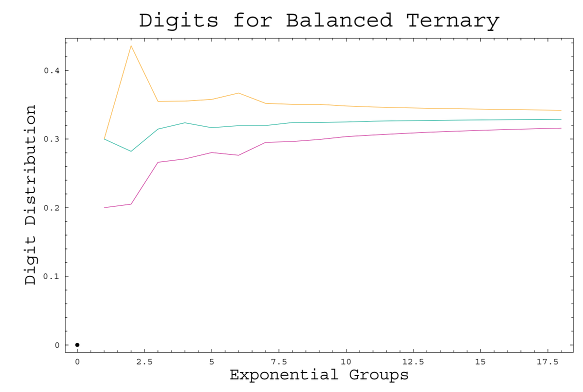
**Table 1: Results for Base 2**

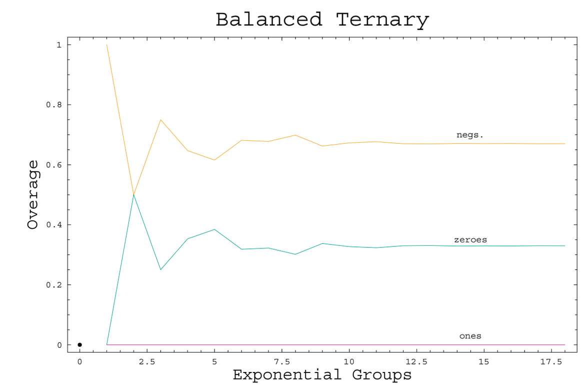
| Start | End | #Primes | Zeroes | Ones | Equals | Evens | Lonely0 | Menorahs |
|---|---|---|---|---|---|---|---|---|
| 4 | 8 | 2 | 1 | 1 | 0 | 1 | 1 | 1 |
| 8 | 16 | 2 | 2 | 2 | 0 | 0 | 2 | 1 |
| 16 | 32 | 5 | 7 | *8 | 0 | 3 | 2 | 2 |
| 32 | 64 | 7 | 13 | *15 | 2 | 2 | 3 | 2 |
| 64 | 128 | 13 | *35 | 30 | 0 | 5 | 0 | 3 |
| 128 | 256 | 23 | *71 | 67 | 4 | 4 | 4 | 4 |
| 256 | 512 | 43 | 147 | *154 | 0 | 23 | 4 | 11 |
| 512 | 1024 | 75 | 298 | *302 | 17 | 27 | 3 | 13 |
| 1024 | 2048 | 137 | *622 | 611 | 0 | 62 | 1 | 21 |
| 2048 | 4096 | 255 | 1270 | *1280 | 28 | 95 | 5 | 46 |
| 4096 | 8192 | 464 | *2558 | 2546 | 0 | 222 | 1 | 64 |
| 8192 | 16384 | 872 | *5257 | 5207 | 189 | 367 | 4 | 109 |
| 16384 | 32768 | 1612 | *10509 | 10447 | 0 | 777 | 0 | 208 |
| 32768 | 65536 | 3030 | *21297 | 21123 | 531 | 1269 | 3 | 375 |
| 65536 | 131072 | 5709 | *42852 | 42783 | 0 | 2910 | 2 | 650 |
| 131072 | 262144 | 10749 | *86258 | 85726 | 1990 | 4859 | 8 | 1171 |
| 262144 | 524288 | 20390 | *173528 | 173102 | 0 | 10140 | 1 | 2068 |
| 524288 | 1048576 | 38635 | *348187 | 347243 | 5747 | 17714 | 11 | 3720 |
| 1048576 | 2097152 | 73586 | *699590 | 698544 | 0 | 36714 | 4 | 6786 |
| 2097152 | 4194304 | 140336 | *1404936 | 1401784 | 23902 | 66020 | 5 | 12410 |
| 4194304 | 8388608 | 268216 | *2818606 | 2813930 | 0 | 133400 | 0 | 22710 |
| 8388608 | 16777216 | 513708 | *5657411 | 5644165 | 76658 | 245959 | 7 | 41461 |
| 16777216 | 33554432 | 985818 | *11345622 | 11328192 | 0 | 493532 | 1 | 76482 |
| 33554432 | 67108864 | 1894120 | *22746823 | 22712057 | 291478 | 916913 | 2 | 141313 |
| 67108864 | 134217728 | 3645744 | *45605127 | 45538473 | 0 | 1822087 | 0 | 261177 |
| 134217728 | 268435456 | 7027290 | *91421299 | 91288241 | 982793 | 3428633 | 1 | 486194 |
| 268435456 | 536870912 | 13561907 | *183206338 | 182965151 | 0 | 6782008 | 5 | 904872 |
| 536870912 | 1073741824 | 26207278 | *367111951 | 366691833 | 3677580 | 12870735 | 4 | 1689754 |
| 1073741824 | 2147483648 | 50697537 | *735525895 | 734702678 | 0 | 25339113 | 0 | 3159326 |
| 2147483648 | 4294967296 | 98182656 | *1473503602 | 1471976078 | 13214719 | 48419194 | 7 | 5927895 |

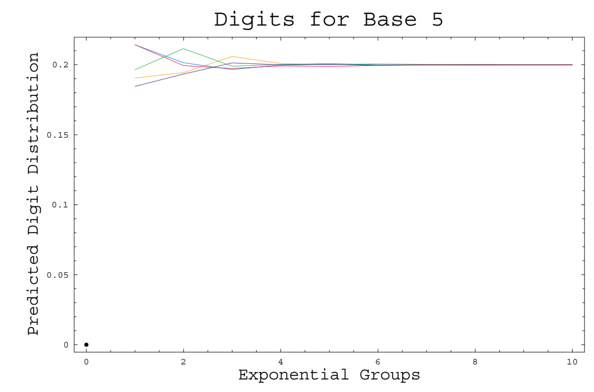
Asterisks denote winning bit.

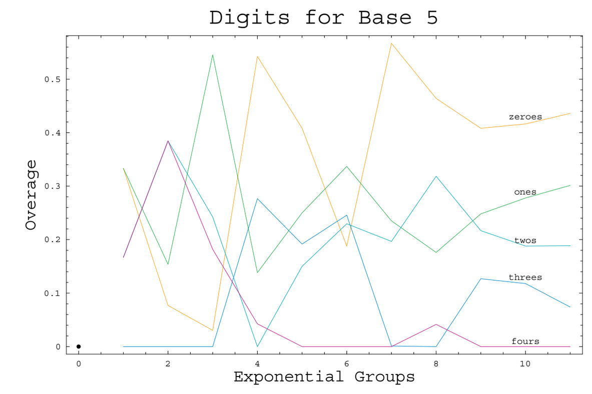



## Conclusions

*Base 2:* The distribution of digits is equal according to Figure 1. However, as prime numbers increased, it became clear the zeroes were slightly more frequent. The ratio of zero bits to one bits in each group remains relatively the same as the primes increase; however, ratios can be misleading. Figure 1 does not show the size of the numbers that comprise the ratios. Though the difference between the number of zero bits and one bits increases, the difference grows slower than the total number of binary bits.

*Base b:* As expected, no matter the base, each digit in the base b representation of primes appears in the same proportion eventually. Surprisingly, when the small differences in number of appearances are studied, the smaller digits consistently appear more often.

## Open Questions

Is the lowest base-bit in fact most prevalent? If so, by what amount are they more prevalent, and what does this tell us about the distribution of the primes? To confirm this or to find further support, more tests could be written to validate the overage principle. More data points for larger prime bases could shed light on the original question.

## Acknowledgements

Many thanks to the School of Science for supporting this project, specifically, Brian Jersky, Cheryl Kelly, and Valerie Burke for their involvement and organization of the Summer Research Program. Finally, I'd like to thank Jim Sauerberg for his unwavering patience, guidance, and willingness to teach.