Are the Digits in a Mersenne Prime Random? A Probability Model

Steven R. Dunbar

Source and Objective
Mersenne Primes
Randomness of Digits
A Probability Model
Conclusion

October 11, 2016
Outline

1. Source and Objective
2. Mersenne Primes
3. Randomness of Digits
4. A Probability Model
5. Conclusion
This talk is my expansion of the blog post: “Strings of Digits in Mersenne Primes”, gottwurfelt.com, Jan 28, 2016
Objective

The objective is to investigate the robustness of a probability model in a simple amusing example. The modeled situation is mathematical, well-understood, and finite so the robustness can be examined. The modeled situation is not surrounded with physical, biological or economic context so that the modeling can be understood in a purely mathematical context.
A Mersenne number is a number of the form

\[ M(n) = 2^n - 1, \]

where \( n \) is an integer. The first few Mersenne numbers are 1, 3, 7, 15, 31, 63, 127, 255, \ldots \) (OEIS A000225).

The number of decimal digits \( d \) in the Mersenne number \( M(n) \) is

\[ d = \lfloor \log(2^n - 1) + 1 \rfloor \approx \lfloor 0.30103n \rfloor + 1. \]
Are the Digits in a Mersenne Prime Random?
A Probability Model

Steven R. Dunbar

Source and Objective
Mersenne Primes
Randomness of Digits
A Probability Model

Conclusion

A *Mersenne prime* is a Mersenne number that is prime. Mersenne primes were named for the French monk Marin Mersenne, who studied these numbers more than 350 years ago. There are only 49 known Mersenne primes.

The most recently discovered (January 7, 2016) Mersenne prime is $M(74,207,281)$, with 22,338,618 digits.
Mersenne Composites

In order for $M(n)$ to be prime, $n$ must itself be prime. This is true since for composite $n$ with factors $r$ and $s$,

$$2^{rs} - 1 = (2^r - 1)(2^{s-1}r + 2^{s-2}r + \cdots + 2^r + 1).$$
Some Sample Mersenne Primes

<table>
<thead>
<tr>
<th>Order</th>
<th>( p )</th>
<th>Digits</th>
<th>( M(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>89</td>
<td>27</td>
<td>618970019642690137449562111</td>
</tr>
<tr>
<td>11</td>
<td>107</td>
<td>33</td>
<td>16225927682921336339157410288127</td>
</tr>
<tr>
<td>12</td>
<td>127</td>
<td>39</td>
<td>17014118346046923173168730371127</td>
</tr>
<tr>
<td>25</td>
<td>21701</td>
<td>6533</td>
<td></td>
</tr>
</tbody>
</table>
Are digits of Mersenne primes random?

Not in base-2!

$$M(p) = 111 \ldots 11$$

But what about the distribution of digits in base-10?
Are the Digits in a Mersenne Prime Random? A Probability Model

Steven R. Dunbar

Source and Objective
Mersenne Primes
Randomness of Digits
A Probability Model
Conclusion

Distribution of Digits

$M_{25} = M(21701)$ has 6533 digits, expect about 650 of each digit.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>657</td>
<td>703</td>
<td>678</td>
<td>682</td>
<td>596</td>
<td>667</td>
<td>655</td>
<td>627</td>
<td>616</td>
<td>652</td>
</tr>
</tbody>
</table>

> chisq.test(digitCount)
Chi-squared test for given probabilities
data: digitCount
X-squared = 14.505, df = 9, p-value = 0.1055

So we fail to reject the null hypothesis that the digits are random (uniformly distributed).
Distribution of Pairs

\[ M_{25} = M(21701) \text{ has 6532 2-digit sequences, expect about 65 of each pair.} \]

<table>
<thead>
<tr>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>...</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>79</td>
<td>65</td>
<td>72</td>
<td>49</td>
<td>69</td>
<td>62</td>
<td>55</td>
<td>64</td>
<td>82</td>
<td>...</td>
<td>41</td>
</tr>
</tbody>
</table>

\[
\text{chisquare}(\text{bidigitCounter}) \\
(114.90753214941824, 0.13096183796521524)
\]

So again we fail to reject the null hypothesis that the digits are random (uniformly distributed).
Are the Digits in a Mersenne Prime Random? A Probability Model

Steven R. Dunbar

Source and Objective
Mersenne Primes
Randomness of Digits
A Probability Model
Conclusion

\[ M_{25} = M(21701) \text{ has 6531 triples, expect about 6 of each.} \]

\[
\begin{array}{cccccccccccc}
000 & 001 & 002 & 003 & 004 & 005 & 006 & 007 & 008 & 009 \\
12 & 7 & 6 & 6 & 3 & 3 & 5 & 9 & 5 & 4 \\
\end{array}
\]

\[ \text{chisquare(tridigitCounter)} \]
\[ (1048.084368657174, 0.13665363987944268) \]

Still we fail to reject the null hypothesis that the digits are random (uniformly distributed).
Distribution of Quadruples

Of the 10,000 possible 4-tuples, only 4,777 are seen in \( M_{25} = M(21701) \). No longer uniformly distributed.

Now we see what is happening: The number of spaces available to hold \( n \)-tuples is \( 6533 - n + 1 \) but the number of possible \( n \)-tuples is \( 10^n \).

For large enough \( n \), there are more possible \( n \)-tuples than available \( n \)-tuples in \( M(p) \). Not all possible \( n \)-digit tuples will be represented.
Probability of a tuple

For example, let a random 4-tuple be specified. Assume the probability of seeing this string in the first 4 places is $10^{-4}$.

Ignore:

- 0 can’t be in leading position,
- only have odd digits in last position
- possibility of something like Benford’s Law
Also we have some information about

$$\mathbb{P}[abcd \text{ in positions } 2,3,4,5|abcd \text{ in positions } 1,2,3,4]$$

So no independence!

*Assume* that the probability of seeing this specified 4-tuple in *any* 4 consecutive digits is independent of the probability of seeing it in *any other* 4 consecutive digits. (This is a false assumption! )
Let the number of digits in $M(p)$ be $d$. Then using the independence assumption the probability distribution of the number of appearances of this $n$-tuple is Binomial($d - n + 1, 10^{-n}$).

Since $d - n + 1 \approx d$ and the probability is small, use the Poisson approximation:

$$\text{Binomial}(d - n + 1, 10^{-n}) \approx \text{Poisson}(d/10^n)$$
The probability of \textit{never} seeing this 4-tuple is then 
\[ P[k = 0] \approx e^{-d/10^n}, \] 
so the probability of encountering this 4-tuple at least once in \( M(p) \) is then 
\[ 1 - e^{-d/10^n}. \]

For \( d = 6533, \ n = 4 \) the probability is 0.47967.
Another Binomial Model

Suppose that the probability of seeing an $n$-tuple is $M(p)$ is $\pi$. Then a binomial model for the number of successes (seeing an $n$-tuple in $M(p)$) among the $10^n$ trials is:

$$\text{Binomial}(10^n, \pi)$$

We expect $10^n\pi$ successes with a variance of $\sqrt{10^n\pi(1 - \pi)}$.

This means we expect accuracy in the first half of the digits of $10^n\pi$. 
Example

Prime $p = 74,207,281$.

d $\approx 0.30103p \approx 22,338,551$

$(1 - \exp(-d/10^7)) \cdot 10^7 \approx 8928853$

$(1 - \exp(-d/10^8)) \cdot 10^8 \approx 20019354$

The digits of $M(74,207,281)$ contain exactly 8,929,592 distinct seven-digit substrings, and exactly 20,021,565 distinct eight-digit substrings.
The digits of the Mersenne Primes are not random, but they act like they are. In fact, the distribution of small tuples of digits can be effectively modeled as though they are random with appropriate distributions.

The model is robust even though the assumptions of constant probability and independence are not satisfied.
Consider the model in a context of a constructed random sequence of digits, where it is already known that the sequence is truly random.

Consider a $100 \times 100$ Markov Chain transition matrix of probability of going from one 2-tuple to the next adjacent 2-tuple in the random string. Investigate the mixing properties of this transition matrix in order to put better bounds on the independence hypothesis.