Investigating the relationship between dice outcomes and Benford's law

Mathematics AA HL

20 pages

1. Introduction

I came across Benford's law when reading about lottery tickets. I discovered that the probability of obtaining a number starting with 1 will be different than that of one starting with 2 and so on... I was immediately fascinated as it described a phenomenon that did not seem intuitive to me. To explore my interest, I investigated more on the distribution of first digits for random variables like phone numbers, social security numbers, Zip Codes… I found that in some cases, the data did follow Benford's law, and in others it did not. This led me to wonder whether or not I could determine under which conditions a random variable follows Benford's law. In this way, I selected what would be the object of this investigation: dice outcomes.

This paper will determine if the different ways of gathering the results of 3000 dice rolled impacts whether or not they follow Benford's law. To reproduce the experience of rolling die, I will use the digital simulation *Dice roll simulator*[1]. A total of 3000 6-sided fair die will be thrown. The sample space of the experiment will be defined:

$$S = \{1,2,3,4,5,6\}$$

Each dice throw will constitute an independent event and will therefore not affect the outcome of the following throw. Additionally, because the dice are fair, there will be an equal probability of rolling each outcome.

The data of the simulation will be gathered in three different ways as follows below.

Method 1: the outcomes of the first 300 dice thrown will be recorded.

Method 2: the outcomes of the 3000 rolls will be divided in 200 groups each including the

[1] "Dice Roll Simulator: Random Rolls with Sum, Average, and Product." *Dice Roll Simulator*, www.free-online-calculator-use.com/dice-roll-simulator.html.

results of 15 consecutive dice throws. The sum of the results of each group will be calculated. Method 3: the dice results will be divided in 200 groups each including the results of 15 consecutive dice throws. The product of the results of each group will be calculated.

The results obtained with each method will then be analysed to establish whether or not they follow Benford's law.

It is relevant to note that the procedure followed in the experiment has limitations. On one hand, the results of the experiment will not be completely random as we are using a simulation. The latter is programmed with algorithms that remove the "chance" from the dice result. On the other, the fact that only 3000 trials will be performed will potentially decrease the accuracy of the experiment, as will later be seen when analysing the results obtain when following methods 1, 2 and 3.

2. Background information

Benford's law, also known as the First Digit Law, is a law that concerns the leading digits of the numbers found in large real-world data sets[2]. It establishes that the frequency of these digits does not follow a uniform distribution. It decreases as leading digits increase from 1 to 9. This law has a series of applications, ranging from detecting fraud in financial statements to describing population distribution in countries and election results. It is given by:

$$P(x) = \log_{10}(x+1) - \log_{10}(x) = \log_{10}\left(1 + \frac{1}{x}\right) \quad (1)$$

Where x ($1 \leq x \leq 9$) is the leading digit. Hence, the probability of a leading digit being 1 would be calculated by replacing the x with 1 as follows:

---

[2] "Benford's Law." *Brilliant Math & Science Wiki*, brilliant.org/wiki/Benford's-law/.

$$P(x = 1) = \log_{10}\left(1 + \frac{1}{1}\right) = 0.301$$

In the same way, we shall use Formula (1) to obtain the probabilities of the remaining leading digits. The results are shown in Table 1.

Leading digit (x)	$P(X = x)$ Benford's law.
1	0.301
2	0.176
3	0.125
4	0.097
5	0.079
6	0.067
7	0.058
8	0.051
9	0.046

Table 1, Probability of a leading digit x according to Benford's Law

Hereunder, I will show and develop the results obtained from Method 1, Method 2 and Method 3 to determine whether or not they follow Benford's law. The first will be explored in depth and the procedure followed will be applied to the second and third.

3. Results

3.1 Method 1

For this first way of gathering data the outcomes of each die of the first 300 throws have been recorded (as a subset of the total population). The frequency of each outcome can be seen on the following chart (View Figure 1) .

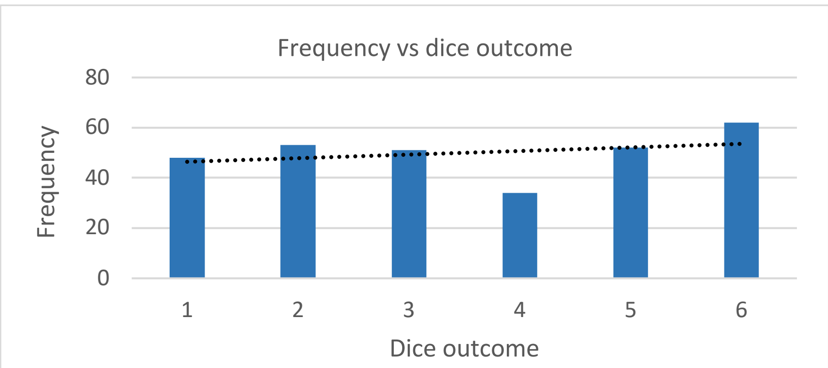


Figure 1, Frequency vs Die outcome, Histogram

As can be seen on Figure 1 there is a relatively uniform distribution among all outcomes with the exception of number 4. This may be due to a small sample size or to errors in the software of the simulator as, theoretically, there should be an equal probability of rolling each number since we are rolling 6-sided fair die.

We can summarize the results obtained in terms of leading digits which will later enable us to compare the data gathered to the first digit law. To do so we will calculate the probability of rolling a leading digit x by using the following formula:

$$P(X = x) = \frac{f}{n}, 0 \leq P(X = x) \leq 1 \quad (2)$$

Where, $P(X = x)$ is the probability of rolling an outcome starting with a leading digit x, f is the frequency of the first digit of each dice result and n is the number of dice rolled.

For example, the probability of rolling an outcome with leading digit 1 will be calculated:

$$P(X = x) = \frac{48}{300} = 0.16$$

In the same way, we shall use Formula 2 for the rest of the leading digits. The results of the calculations, rounded to 3 s.f as given by Benford's law, can be seen on *Table 2*:

Leading digit (x)	$P(X = x)$ Method 1
1	0.16
2	0.177
3	0.17
4	0.113
5	0.173
6	0.207
7	0
8	0
9	0

Table 3, Probability of rolling an outcome whose first digit is x

3.1.2 Analysis

To determine whether or not the data gathered with Method 1 follows Benford’s law we will compare and contrast P(X = x) in Method 1 with P(X = x) in Benford’s law.

3.1.2.1Graphical analysis

A histogram can be plotted with P(X=x) in Method 1 and P(X=x) in Benford’s law with the data in Tables 1 and 2. View Table 3 and Figure 2.

Leading digit (x)	$P(X = x)$ Method 1	$P(X = x)$ Benford’s law.
1	0.16	0.301
2	0.177	0.176
3	0.17	0.125
4	0.113	0.097
5	0.173	0.079
6	0.207	0.067
7	0	0.058
8	0	0.051
9	0	0.046

Table 4, Probability of a leading digit x in Method 1 and Benford's Law

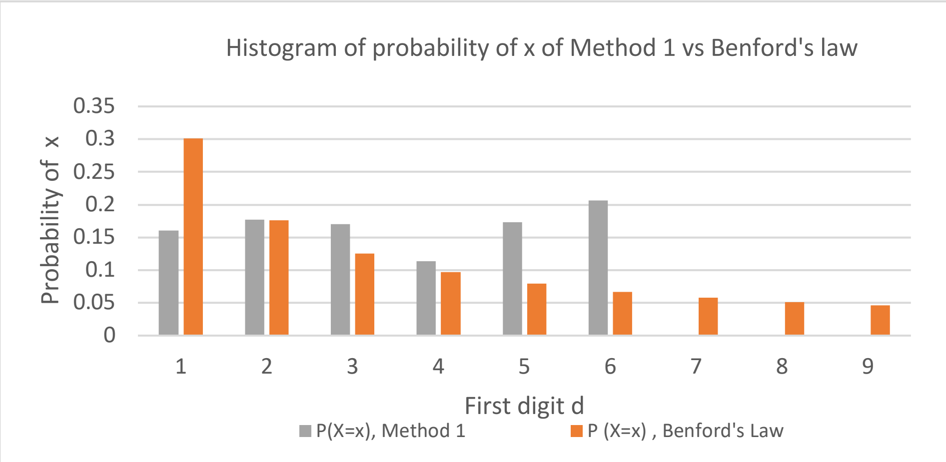


Figure 2, Probability of a leading digit x in Method 1 and Benford's law, Histogram

On one hand, values in P(X = x) in Method 1 only spread form first digits 1 to 6 (beyond this point P(X = x) is 0 and we can consider it to be insignificant). This is to be expected as the dice rolled are 6-sided. On the contrary, values given by Benford’s law spread from first digit 1 to 9. As shown on Figure 2 the first digit distribution for values starting with 2, 3, and 4 are similar for P(X = x) in Method 1 and P(X = x) in Benford's law. Conversely, those for 1, 5, 6 are very different. We can infer from the above that there is a slight relationship between both variables.

3.1.2.2 Statistical analysis

To support this statement, we can compare the statistical measures of the data in Table 2 to those given by Benford's law (these have been obtained from Pampa University[3]). We will calculate the measures of central tendency (mode, median, expected value), spread (quartiles, standard deviation and variance) and skewness (Pearson's coefficient of skewness).

Measures of central tendency

- Mode: it is the value with the highest probability. As shown on Table 2 it is equal to 6.
- Median: It is the second quartile. It is calculated: P(X=x) = 0.5 this occurs when x is equal to 3.
- Expected Value. It is mathematically expressed:

$$E(X) = \sum xp \quad (3)$$

Where $E(X)$ is the expected value, x is the leading digit and p the probability of obtaining a number whose first digit is x. Substituting with the values found in Table 2 these are my results:

$$E(x) = (1 \cdot 0.16) + (2 \cdot 0.1766667) + (3 \cdot 0.17) \ldots + (9 \cdot 0) = 3.58333$$

[3] Cerani, Sebastián, and Manuel Olivera. "Estadística, Probabilidad y Estadística II." Universidad Nacional De La Pampa, 2015.

Measures of spread

- Lower quartile (Q_1): P(X=x) = 0.25, this occurs when x is 2
- Upper quartile(Q_3): P (X=x) = 0.75, this occurs when x is 5
- Interquartile range (IQR)

$$IQR = Q_3 - Q_1 \quad (4)$$

Where Q_3 is the upper quartile and Q_1 is the lower quartile. Substituting each value, the result is the following:

$$IQR = 5 - 2 = 3$$

- Standard deviation (σ):

$$\sigma = \sqrt{x^2 P\ (X = x) - \mu^2} \quad (5)$$

Where x is the leading digit, P(X=x) the probability of obtaining x and μ the expected value. Substituting with the values obtained in Table 2 and with Formula 3 the results are:

$$\sigma = \sqrt{(1^2 \cdot 0.16) + (2^2 \cdot 0.1766667) + (3^2 \cdot 0.17) + (4^2 \cdot 0.1133333) \ldots + (9^2 \cdot 0) - 3.58333\ ^2}$$

$$= 1.77287$$

- Variance, $VAR\ (X)$

$$VAR(X) = \sum x^2 P(X = x) - \mu^2 \quad (6)$$

Where $VAR(X)$ is the variance, x is the leading digit, P(X=x) the probability of obtaining x and μ the expected value. Substituting with the values obtained in Table 2 and with Formula 3 the results are:

$$VAR(X) = (1^2 \cdot 0.16) + (2^2 \cdot 0.1766667) + (3^2 \cdot 0.17) + (4^2 \cdot 0.1133333) \ldots + (9^2 \cdot 0) - 3.58333\ ^2$$

$$= 3.14307$$

Measures of Skewness

- Pearson's coefficient of skewness: it is calculated:

$$sk_1 = \frac{\mu - Mo}{\sigma} \quad (7)$$

Where sk_1 is Pearson's coefficient of skewness, μ is the expected value, Mo is the mode, σ is the standard deviation. This coefficient will allow us to numerically describe and determine the way in which is the data is distributed and whether it fits a specific type. Substituting with the values found with Formulas 3 and 4 and our data the results are:

$$sk_1 = \frac{3.58333 - 3}{1.77287} = 0.329031$$

Having analysed the data, we can proceed to establish whether or not the data in Method 1 follows Benford's Law

3.1.3 Interpretation

We can display and compare the values obtained in the statistical analysis with those established by Benford's Law as can be seen below in Table 3. The values have been rounded to the same number of significant figures that are given by the statistical measures of Benford's law.

Statistical Measures	Method 1	Benford's Law
Mode (*Mo)*	3	1
Q_1	2	1
Median (*Md)*	3	3
Q_3	5	5
IQR	3	4
Expected value(EX*)*	3.583330	3.440237
Standard deviation (σ)	1.773	2.461
Variance (σ^2)	3.14307	6.05652
Pearson's coefficient of skewness (sk_1)	0.3290	0.7958

Table 5, Statistical measures of data in Method 1 vs Statistical measures of Benford's Law.

From Table 4, we can make the following observations regarding the measures of central tendency, spread and skewness:

1. Measures of central tendency: firstly, the mode or most frequent value is 6 as opposed to 1 as established by Benford's law. Secondly, both the median of the data and the expected value are equal to or approximate respectively to the values established by Benford's law.
2. Measures of spread: both the third and first quartile approximate to those of Benford's law. Both the standard deviation and the variance are roughly half of those established by Benford's law.
3. Measure of skewness: As given by Pearson's coefficient of skewness the data is positively skewed ($sk_1 > 0$) , although the asymmetry is small as the coefficient's value is close to 0. Conversely, in Benford's law, the first digit distribution must be strongly positively skewed.

It appears that there is a slight parallelism between the probability of a first digit x we obtained and that established by Benford's Law. Some values are equal to or close to those established by the first digit law. Others show a clear relationship (like for example being equal to ½ of those of the law). The data of Method 1 slightly follows Benford's law.

To conclude, we can use Pearson's correlation coefficient to mathematically confirm this. The latter is used to measure the strength of a linear association between two variables[4]. It is mathematically expressed:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \quad (8)$$

[4] *Pearson Correlation Coefficient Calculator*, www.socscistatistics.com/tests/pearson/.

Where x is the first variable, $\bar{x}$ is the mean of the values given by x, y is the second variable and $\bar{y}$ is the mean of the values given by y. We shall substitute x and y with P(X = x) of Method 1 and P(X = x) of Benford's Law respectively. The means for each will be calculated:

$$\bar{x} = \frac{\sum x}{n} \quad (9)$$

Where, $\bar{x}$ is the mean, x represents the values given by a certain variable and n is the number of values that there are. Substituting for the two variables:

$$\bar{x} = \frac{\sum x}{n} = \frac{1}{9}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{1}{9}$$

Given the values of the means we can proceed to calculate the value of r. The results obtained, rounded to 3 s.f can be viewed below:

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
0.16	0.301	0.049	0.002	0.19	0.036	0.009
0.177	0.176	0.066	0.004	0.065	0.004	0.004
0.17	0.125	0.059	0.003	0.014	0	0.001
0.113	0.097	0.002	0	-0.014	0	0
0.173	0.079	0.062	0.004	-0.032	0.001	-0.002
0.207	0.067	0.096	0.009	-0.044	0.002	-0.004
0	0.058	-0.111	0.012	-0.053	0.003	0.006
0	0.051	-0.111	0.012	-0.06	0.004	0.007
0	0.046	-0.111	0.012	-0.065	0.004	0.007
Sum		$\sum(x-\bar{x})$ = 0.001	$\sum(x-\bar{x})^2$ =0.060	$\sum(y-\bar{y}) \approx$ 0.001	$\sum(y-\bar{y})^2$=0.054	$\sum(x-\bar{x})(y-\bar{y})$ = 0.028

Table 6, Pearson's correlation coefficient calculations

Replacing these values in the formula (9):

$$r = \frac{0.028}{\sqrt{0.060 \cdot 0.054}} = 0.492$$

The final value of r, rounded to 3 s.f, 0.492 indicates a weak positive correlation. It must be noted it approximates of being a moderate correlation (r = 0.5). Therefore, we can state that whilst the results obtained do exhibit a relationship with Benford's law, they do comply with it to a high degree. These findings are consistent throughout the analysis and the interpretation of Method 1's data. Hence, it is shown that the results of dice thrown if gathered as established in Method 1, slightly follow the first digit law.

3.2 Method 2

For this second method of gathering the results of the 3000 dice rolled, the outcomes of the dice have been divided into 200 groups. Each containing 15 consecutive dice results that have been added up together (View Appendix A). The frequency of each sum result has been recorded as follows.

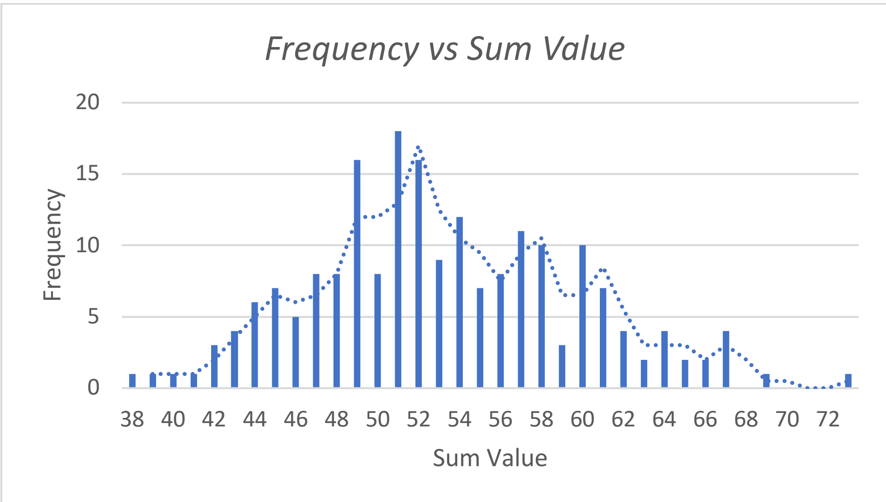


Figure 3, Frequency vs Sum Value, Histogram

The data recorded appears to be normally distributed. We can state this for two reasons. The first being the rough bell-like shape the frequency vs sum graph and the histogram of the first digit present. It is characteristic of a normal distribution. The second being the Central Limit Theorem or CLT (probability theory), according to which summing independent random variables will tend towards a normal distribution. We will later verify whether or not this is so by using Pearson's coefficient of skewness.

We can summarize the results obtained in terms of leading digits. We will calculate the probability of obtaining a sum whose first digit is x by using formula (2) in the same way we have in Method 1. In this case, we will substitute with the values of Figure 3. The results of the calculations, rounded to 3 s.f as given by Benford's law, can be seen on *Table 6*:

First digit d	$P(X = x)$ Method 2
1	0
2	0
3	0.01
4	0.295
5	0.51
6	0.18
7	0.005
8	0
9	0

Table 7, Probability of obtaining a sum whose first digit is x

3.2.2 Analysis

3.2.2.1 Graphical analysis

A histogram can be plotted with P(X=x) in Method 2 and P(X=x) in Benford's law with the data in Tables 1 and 6. View Table 7 and Figure 2.

Leading digit (x)	$P(X = x)$ Method 1	$P(X = d)$ Benford's law.
1	0	0.301
2	0	0.176
3	0.01	0.125
4	0.295	0.097
5	0.51	0.079
6	0.18	0.067
7	0.005	0.058
8	0	0.051
9	0	0.046

Table 8, Probability of a leading digit x in Method 2 and Benford's Law

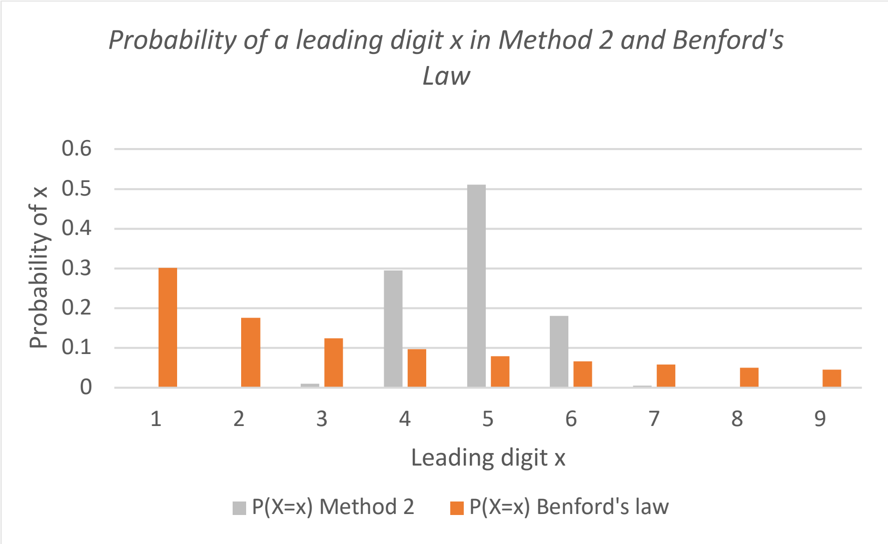


Figure 4, Probability of a leading digit x in Method 2 and Benford's Law, Histogram

We can observe that the distribution of the probabilities is very different from that of Benford's Law. Values in Method 2 are only spread among first digits 3, 4, 5, 6 and 7. On the contrary, values given by Benford's law spread from first digit 1 to 9. This creates a substantial difference between both data sets. Furthermore, none of the probabilities of Method 2 are similar to those of the first digit law. From the above, we can state that there seems to be little relationship between both data sets.

3.1.2.2 Statistical analysis

To support this statement, we can compare the statistical measures of the data in Table 6 to those given by Benford's law. We will calculate the measures of central tendency, spread and skewness by substituting the corresponding values (Table 6) in formula's (3), (4), (5), (6) and (7). The results can be viewed in the following section (Table 8) directly compared to those of the first digit law. The values have been rounded to the same number of significant figures that are given by Benford's law.

3.2.3 Interpretation

Statistical Measures	Method 2	Benford's Law
Mode (*Mo)*	5	1
Q_1	4	1
Median (*Md)*	5	3
Q_3	5	5
IQR	1	4
Expected value E(X*)*	4.875	3.440237
Standard deviation (σ)	0.721	2.461
Variance VAR(X)	0.519408	6.05652
Pearson's coefficient of skewness (sk_1)	-0.1734	0.7958

Table 9, Statistical measures of data in Method 2 vs Statistical measures of Benford's Law.

From the table above we can make the following observations regarding the measures of central tendency, spread, and skewness:

1. Measures of central tendency: they are substantially different between Method 2 and Benford's law.
2. Measures of spread: Both the first and the third quartile coincide. The median is different. This can be explained by noting that the sum results of method 2 only range from first digit 3 to 7 as explained in 3.2.2.1. The standard deviation and variance are both smaller than in the first digit law as most values are concentrated at the centre of the distribution (approximately 50% of the values are within one standard deviation away from the expected value, 88% two and 100% three, note these brackets are similar to those of a normal distribution).
3. Measure of skewness: the coefficient of skewness approximates to zero (sk_1= -0.1734) which is characteristic of normal distributions (they are symmetrical). This reinforces comments made on distribution. Therefore, we can finally establish that both the data itself is normally distributed. To follow Benford's law the skewness it should be positive and stronger.

It appears that there is little to no correlation between the probability of a first digit x in the sum and that established by Benford's Law. To conclude, we can use Pearson's correlation coefficient to verify this statement. We shall follow the steps established in Method 1, this time using $P(X = x)$ in Method 2 as the first variable and $P(X = x)$ in Benford's law as the second. We will replace the corresponding values in formulas (8) and (9). The resulting coefficient rounded to 3 s.f (View Appendix B) will be $r = -0.2371$, which indicates a very weak negative correlation. That is to say, that gathering dice results and adding them up will not follow the First digit law. As with Method 1, the spread of the values gathered is too small and the distribution of data is incompatible with Benford's law.

3.3 Method 3

For this second method of gathering the results of the 3000 dice rolled, the outcomes of the dice have been divided into 200 groups. Each containing 15 consecutive dice results that have been multiplied together (View Appendix C). The frequency of each product can be seen below:

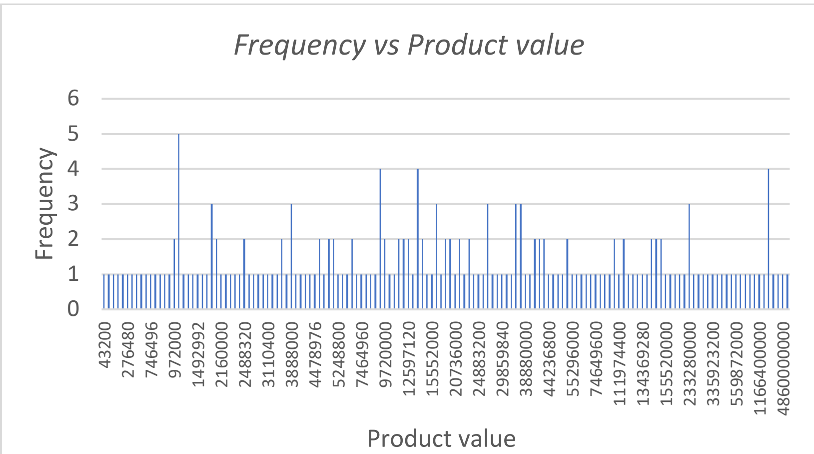


Figure 5, Frequency vs Product Value, Histogram 1

It is difficult to make comments on the distribution of the product values in *Figure 5* as there are large intervals between numbers and the frequencies do not immediately show a clear shape. We can attempt to plot a histogram this time plotting the product value frequencies in intervals to make the distribution of data clearer.

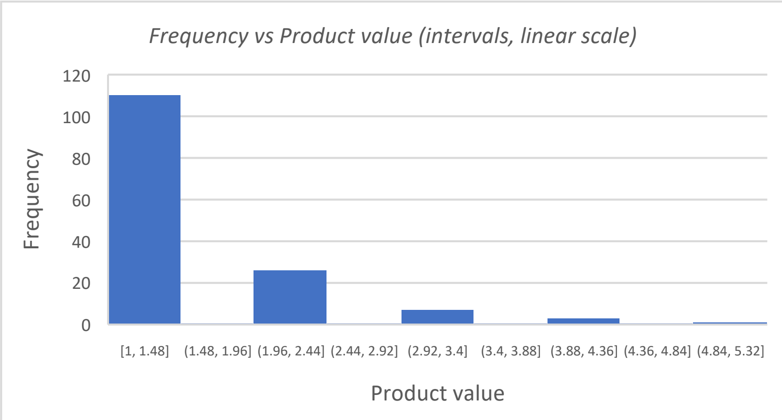


Figure 6, Frequency vs Product values, Histogram 2

On Figure 6 we can observe that the data is positively skewed and creates a log-like shape (as shown by the dark trendline) on a linear scale. We can also plot the data on a logarithmic scale and observe how the data behaves (view Figure 7).

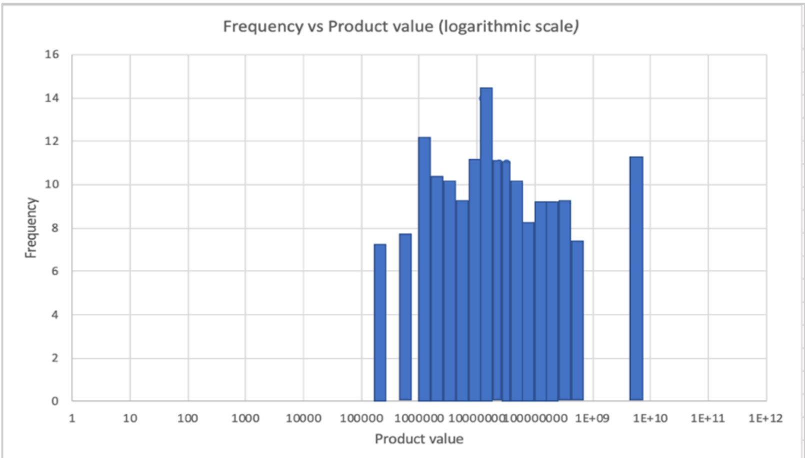


Figure 7, Frequency vs Product value, Histogram 3

On a logarithmic scale it would appear that the data creates the rough bell-like shape characteristic of normal distribution. We can thereby begin to consider that the distribution of the data is lognormal. From a theory supported standpoint this should be so as the Multiplicative Central Limit Theorem (MTLT, an extension of CLT of probability theory) states that multiplying independent random variables (like dice outcomes) will tend towards lognormal distribution. We will later verify this mathematically through Pearson's coefficient of skewness as we did in Methods 1 and 2.

We can summarize the results obtained in terms of leading digits. We will calculate the probability of obtaining a product whose first digit is x by using formula (2) in the same way we have done throughout. In this case, we will substitute with the values of Figure 5. The results of the calculations, rounded to 3 s.f as given by Benford's law, can be seen on *Table 9*:

First digit d	$P(X = x)$ Method 3
1	0.345
2	0.185
3	0.15
4	0.095
5	0.055
6	0.025
7	0.045
8	0.035
9	0.065

Table 10, Probability of obtaining a product whose first digit is x

3.3.2 Analysis

3.3.2.1 Graphical analysis

A histogram can be plotted with P(X = x) in Method 2 and P(X = x) in Benford's law with the data in Tables 1 and 9. View Table 10 and Figure 8.

Leading digit (x)	$P(X = x)$ Method 1	$P(X = x)$ Benford's law.
1	0.345	0.301
2	0.185	0.176
3	0.15	0.125
4	0.095	0.097
5	0.055	0.079
6	0.025	0.067
7	0.045	0.058
8	0.035	0.051
9	0.065	0.046

Table 11, Probability of a leading digit x in Method 2 and Benford's Law

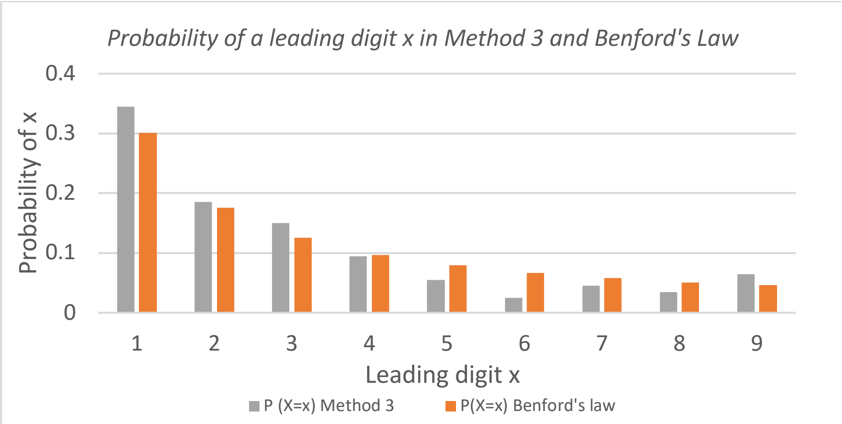


Figure 8,Probability of a leading digit x in Method 2 and Benford's Law, Histogram

It appears that the first digit distribution of the product values and that of Benford's law is very similar. Particularly this is so for digits 1, 2, 3 and 4. From this point onwards the data seems to differ more significantly although there is still a visible parallelism. This may be due to the limited size of the population (with more data it would likely become more accurate). Furthermore, Method 3 and Benford's Law display the same range of leading digits which contributes to the similarity. From the above, we can state there is a relationship between both data sets.

3.3.2.2 Statistical analysis

To support this statement, we can compare the statistical measures of the data in Table 9 to those given by Benford's law. We will calculate the measures of central tendency, spread skewness by substituting the corresponding values (Table 9) in formula's (3), (4), (5), (6) and (7). The results can be viewed in the following section (Table 11) directly compared to those of the first digit law. The values have been rounded to the same number of significant figures that are given by Benford's law.

Statistical Measures	Method 3	Benford's Law
Mode (*Mo*)	1	1
Q_1	1	1
Median (*Md*)	2	3
Q_3	4	5
IQR	3	4
Expected value E(X)	3.15	3.440237
Standard deviation (σ)	2.453	2.461
Variance VAR(X)	6.0175	6.05652
Pearson's coefficient of skewness (sk_1)	0.8765	0.7958

Table 12, Statistical measures of data in Method 3 vs Statistical measures of Benford's Law

3.3.3.Interpretation

From Table 11 we can make the following observations:

1. Measures of central tendency: they are very close or equal to those of Benford's law.
2. Measures of spread: they quartiles are virtually identical, either being 1 unit lower or equal. The standard deviations differ by 0.008 and the variance by 0.039. They are thus also virtually identical.
3. Measure of skewness: the data in Method 3 data is strongly positively skewed, which coincides with the lognormal distribution hypothesis formulated previously. Furthermore, it has a distribution of first digits that is shaped in the same fashion that that of Benford's law (it is simply stronger).

It appears that there is strong correlation between the probability of a first digit x in the product of dice outcomes and that established by Benford's Law. To conclude, we can use Pearson's correlation coefficient to verify this statement. We shall follow the steps established in Method 1, this time using P(X = x) in Method 3 as the first variable and P(X = x) in Benford's law as the second. We will replace the corresponding values in formulas (8) and (9). The resulting coefficient rounded to 3 s.f (View appendix D) will be $r = 0.982$, which indicates a very strong, almost perfect positive correlation. That is to say, that gathering dice results and multiplying them will produce a data set that follows the First digit law.

4. Conclusion

Out of the three methods tested for gathering dice results only Method 3 follows the first digit law. In Methods 1 and 2, the distribution of data was uniform and normal respectively. This ultimately limited the spread of the first digits to a smaller range than that of the first digit law. From this we can extrapolate that data sets that conform to these distributions will not follow Benford's law. Conversely, Method 3 followed a log-normal distribution and therefore had a wider spread and range of leading digits. This coincides with theory as: *Benford's law applies to data sets whose distributions conform to a lognormal distribution and whose standard deviation exceeds 1.2. Data sets that are likely to satisfy this criterion will:*

- *Have only positive values.*
- *Have a unimodal distribution whose modal value is not zero.*
- *Have a positively skewed distribution in which the median is no more than half of the mean*[5].

If we observe Table 11 one last time, we can observe that our data fulfils all of these criteria. In this way, our findings throughout this paper rightfully coincide with theoretical predictions. Hence, in conclusion we can affirm that throwing dice can lead to Benford's Law if and only if the results are gathered in such a way that they create a log-normal distribution.

[5] Scott, P. D., and M. Fasli. "Benford's Law: An Empirical Investigation and a Novel Explanation." University of Essex.

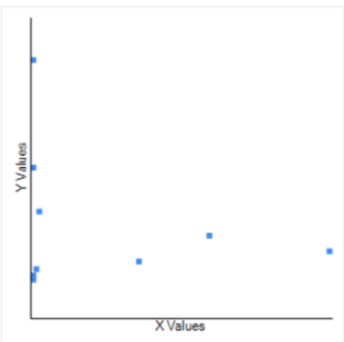
Appendices

Appendix A: Method 2 Raw Results

Sum of the outcomes of each trial							
46	58	54	50	55	54	64	61
49	54	52	49	51	44	49	45
53	48	60	50	53	50	65	51
63	42	52	60	49	49	48	47
53	66	51	56	47	63	58	43
55	62	57	58	38	50	48	48
39	49	53	60	45	41	64	51
67	46	54	56	49	53	54	61
44	61	52	47	52	57	56	60
52	58	61	51	52	58	52	58
57	48	73	52	57	51	40	44
57	64	50	50	49	58	47	56
67	53	59	45	49	51	51	60
46	62	55	58	61	52	46	52
51	48	57	49	51	61	51	44
67	48	47	54	53	56	53	58
51	60	49	45	55	44	58	55
62	61	51	57	43	54	49	54
60	60	51	46	48	53	57	64
58	45	59	60	66	69	44	52
45	54	43	56	55	45	52	42
49	60	51	54	54	57	47	54
62	51	55	57	49	49	51	52
65	59	47	50	52	56	42	49
47	67	52	43	52	51	50	56

Appendix B, Pearson's Correlation Coefficient, Method 2.

X Values	Y Values	$X - M_x$	$Y - M_y$	$(X - M_x)^2$	$(Y - M_y)^2$	$(X - M_x)(Y - M_y)$
0	0.301	-0.111	0.190	0.012	0.036	-0.021
0	0.176	-0.111	0.065	0.012	0.004	-0.007
0.01	0.125	-0.101	0.014	0.010	0.000	-0.001
0.3	0.097	0.189	-0.014	0.036	0.000	-0.003
0.505	0.079	0.394	-0.032	0.155	0.001	-0.013
0.18	0.067	0.069	-0.044	0.005	0.002	-0.003
0.005	0.058	-0.106	-0.053	0.011	0.003	0.006
0	0.051	-0.111	-0.060	0.012	0.004	0.007
0	0.046	-0.111	-0.065	0.012	0.004	0.007
		Mx: 0.111	My: 0.111	Sum: 0.266	Sum: 0.054	Sum: -0.029



Result Details & Calculation

X Values
$\Sigma = 1$
Mean = 0.111
$\Sigma(X - M_x)^2 = SS_x = 0.266$

Y Values
$\Sigma = 1$
Mean = 0.111
$\Sigma(Y - M_y)^2 = SS_y = 0.054$

X and Y Combined
$N = 9$
$\Sigma(X - M_x)(Y - M_y) = -0.029$

R Calculation
$r = \Sigma((X - M_y)(Y - M_x)) / \sqrt{((SS_x)(SS_y))}$

$r = -0.029 / \sqrt{((0.266)(0.054))} = -0.2371$

Meta Numerics (cross-check)
$r = -0.2371$

Key

X: X Values
Y: Y Values
M_x: Mean of X Values
M_y: Mean of Y Values
$X - M_x$ & $Y - M_y$: Deviation scores
$(X - M_x)^2$ & $(Y - M_y)^2$: Deviation Squared
$(X - M_x)(Y - M_y)$: Product of Deviation Scores

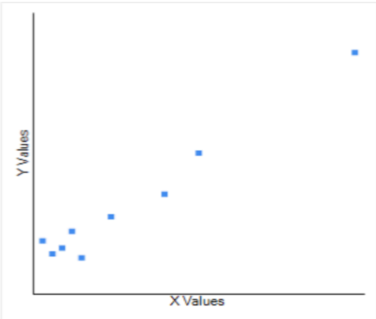
Appendix C, Method 3 raw data.

Product of the outcomes of each trial							
972000	27993600	16588800	9216000	39813120	27648000	1415577600	77760000
3456000	46656000	13996800	4860000	4976640	2211840	3888000	2488320
37324800	12597120	139968000	5184000	55296000	6220800	1166400000	11059200
279936000	972000	16588800	174960000	2400000	7464960	1866240	2764800
10497600	1244160000	39321600	29859840	1866240	233280000	124416000	829440
13996800	382205952	72000000	97200000	46656	4478976	4976640	2160000
96000	4665600	51840000	155520000	864000	43200	839808000	9720000
2015539200	2457600	23328000	45349632	3456000	15552000	34992000	265420800
1866240	233280000	9331200	2916000	18662400	23887872	25194240	143327232
9331200	149299200	186624000	9331200	7200000	41472000	13996800	143327232
82944000	3110400	12441600000	20736000	38880000	14580000	230400	259200
41472000	179159040	18662400	3888000	6220800	69984000	2799360	43200000
1259712000	23328000	233280000	1036800	4665600	11059200	13996800	331776000
2073600	298598400	72900000	69984000	796262400	19906560	1327104	12597120
1259712000	4423680	39813120	14929920	9331200	134369280	12441600	576000
1259712000	3732480	2073600	22394880	25920000	44789760	23040000	149299200
4147200	116640000	3981312	1749600	44236800	1492992	43200000	25920000
335923200	349920000	3888000	34992000	364500	16588800	3317760	22394880
131220000	139968000	12441600	1244160	2592000	34992000	97200000	746496000
194400000	1658880	67184640	129600000	559872000	4860000000	1036800	32400000
1036800	10368000	839808	388800000	136687500	1152000	19906560	388800
3240000	524880000	12960000	9720000	74649600	37324800	4199040	31104000
116640000	8294400	17915904	37324800	2332800	5598720	22394880	14580000
1119744000	111974400	933120	8398080	25920000	51840000	276480	2488320
1036800	1259712000	5248800	384000	24883200	5971968	5184000	79626240

Appendix D, Pearson's Correlation Coefficient, Method 2.

X Values	Y Values
0.345	0.301
0.185	0.176
0.15	0.125
0.095	0.097
0.055	0.079
0.025	0.067
0.045	0.058
0.035	0.051
0.065	0.046

$X - M_x$	$Y - M_y$	$(X - M_x)^2$	$(Y - M_y)^2$	$(X - M_x)(Y - M_y)$
0.234	0.190	0.055	0.036	0.044
0.074	0.065	0.005	0.004	0.005
0.039	0.014	0.002	0.000	0.001
-0.016	-0.014	0.000	0.000	0.000
-0.056	-0.032	0.003	0.001	0.002
-0.086	-0.044	0.007	0.002	0.004
-0.066	-0.053	0.004	0.003	0.004
-0.076	-0.060	0.006	0.004	0.005
-0.046	-0.065	0.002	0.004	0.003
Mx: 0.111	My: 0.111	Sum: 0.085	Sum: 0.054	Sum: 0.067



Result Details & Calculation

X Values
$\Sigma = 1$
Mean = 0.111
$\Sigma(X - M_x)^2 = SS_x = 0.085$

Y Values
$\Sigma = 1$
Mean = 0.111
$\Sigma(Y - M_y)^2 = SS_y = 0.054$

X and Y Combined
$N = 9$
$\Sigma(X - M_x)(Y - M_y) = 0.067$

R Calculation
$r = \Sigma((X - M_y)(Y - M_x)) / \sqrt{((SS_x)(SS_y))}$

$r = 0.067 / \sqrt{((0.085)(0.054))} = 0.9824$

Meta Numerics (cross-check)
r = 0.9824

Key

X: X Values
Y: Y Values
M_x: Mean of X Values
M_y: Mean of Y Values
$X - M_x$ & $Y - M_y$: Deviation scores
$(X - M_x)^2$ & $(Y - M_y)^2$: Deviation Squared
$(X - M_x)(Y - M_y)$: Product of Deviation Scores

References

Benford's Law." *Brilliant Math & Science Wiki*, brilliant.org/wiki/Benford's-law/.

Cerani, Sebastián, and Manuel Olivera. "Estadística, Probabilidad y Estadística II." Universidad Nacional De La Pampa, 2015.

"Dice Roll Simulator: Random Rolls with Sum, Average, and Product." *Dice Roll Simulator*, www.free-online-calculator-use.com/dice-roll-simulator.html.

Pearson Correlation Coefficient Calculator, https://www.socscistatistics.com/tests/pearson/default.aspx

Scott, P. D., and M. Fasli. "Benford's Law: An Empirical Investigation and a Novel Explanation." University of Essex.