

# Benford's Law の統計的解析

2003MM093 佐々 裕典

2003MM097 鈴木 麻峰

2003MM099 鈴木 快斗

指導教員 尾崎 俊治

## 1 はじめに

私達の身の周りには、無限の数が存在している。それは、実生活においてなくてはならないものである。

本論文では、その無限の数において、人為的な意図が加わっていない数値の最初の桁の数値を考えるとする。1桁目の数値を分類した場合、本論文の Benford's Law を知らない人であったならば、1~9の最初の桁の数値は均等に分布していると考えの人が多くであろう。しかし、1~9の数値の分布は、その予想に反して、ある一定の割合に限りなく近くなる。

本論文では、実データやフィボナッチ数列等、多方面から Benford's Law を研究し、論じていく。

## 2 Benford's Law とは

Benford's Law は、1881年にアメリカ人の天文学者ニューカムが、基盤となるものを発見した。それは、対数表の膨大なページを見た際に、数多くの対数が1で始まることに気がついたことである。

ニューカムは、対数のみで数の偏りがあると気がついたが、物理学者のベンフォードは、それを基に世の中に存在する、20,229もの膨大な統計データを用い、それらの最初の桁の数値の分布を調べた [1]。

その結果、人為的な意図が加わっていないデータの最初の桁の数値は、ある一定の割合に近づくことを発見した。その割合を表1に示す。

また、数式的に表すと、最初の桁の数値  $n(n=1\sim 9)$  の割合は  $f(n)$  を用いると (1) 式のように表現される [2]。

$$\begin{aligned} f(n) &= \log_{10}(n+1) - \log_{10}(n) \\ &= \log_{10}\left(\frac{n+1}{n}\right) \\ &= \log_{10}\left(1 + \frac{1}{n}\right) \end{aligned} \quad (1)$$

この数式に、最初の桁の数値  $n$  を代入し計算すると、最初の桁の数値の割合は表1の傾向に近づく。

グラフに表すと図1となる。図1の  $x$  軸は、最初の桁の数値、 $y$  軸はその割合を表す。

表1、図1を見てわかるように、Benford's Law に従ったグラフは、1の割合が1番多くなり、1~4の割合が約70%占めている。また、2, 3, 4と最初の桁数が増加していくにつれて、その割合が減少していることがわかる。

表1 最初の桁の数値の割合

1	30.1%
2	17.6%
3	12.5%
4	9.7%
5	7.9%
6	6.7%
7	5.8%
8	5.1%
9	4.6%

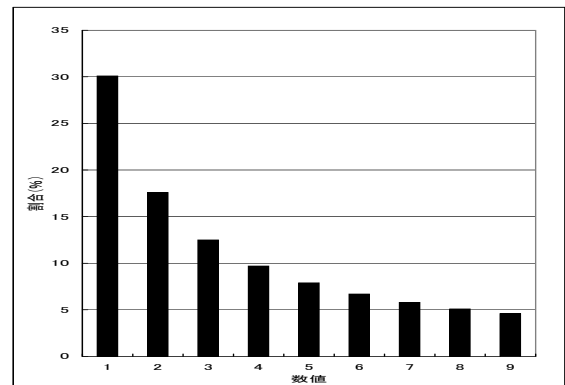


図1 Benford's Law の割合

## 3 実データによる検証

この章では、まず最初に扱う実データについての説明を行う。次にそのデータの解析した結果を表を用いて説明する。最後に得られた結果から考察を行う。3章以降では与えられたデータに対して、Benford's Law に従っているかどうか数学的根拠を得るためにカイ2乗検定 [3] を行う。

### 3.1 扱う実データ

最初に、実データの値が Benford's Law に従うか実際に検証した。

その実データとして、発行済株式数 [4]、日本の各都市人口動態 [5]、日本の各都市上水道給水状況 [5]、日本の各都市の自動車保有車両数 [5]、世界の人口 [6]、世界の面積 [6]、世界の自動車保有台数 [6]、世界の鉄道運送量 [6] を扱った。これらのデータは、ランダムな数値かつ人為的な意図が加わってないと考えられたため扱った。

### 3.2 結果

8つの実データの最初の桁の数値の割合を計算し、Benford's Law に従う結果となるか否か、カイ2乗検定を行った。その検定結果を表2に示す。

カイ2乗検定の結果からわかる事は、各々のデータ単体

表2 カイ2乗検定

	カイ2乗値
発行株式数	0.187
各都市人口	$1.15 \times 10^{-3}$
上下水給水	$9.21 \times 10^{-4}$
日本の自動車保有数	$1.44 \times 10^{-2}$
世界の人口	$6.16 \times 10^{-3}$
世界の面積	$8.22 \times 10^{-3}$
世界の自動車保有数	0.918
世界の鉄道運送量	4.969
平均	$6.49 \times 10^{-7}$

で検定すると、カイ2乗値は大きいですが、8個のデータの平均の割合を出し、カイ2乗検定を行った場合、カイ2乗値は小さい結果となった。

### 3.3 考察

有意水準を95%に設定し、仮説検定を行う。帰無仮説として、『データの割合がBenford's Lawの割合に従わない』と仮定する。95%点のカイ2乗値は、 $3.9 \times 10^{-3}$ であり、この数値より小さいデータは、各都市人口、上下水給水であり、それ以外のデータは、95%点のカイ2乗値より大きい数値となる。ゆえに、各都市人口、上下水給水は、有意水準95%より大きい確率となり、その他のデータは、有意水準95%より小さい確率となる。

有意水準を95%に設定したために、各都市人口、上下水給水のデータにおいては、帰無仮説が棄却され、対立仮説である『データの割合がBenford's Lawに従う』が採択される。その他のデータは、帰無仮説が採択される結果となった。

データ一つ一つにおいてカイ2乗検定を行った場合、有意水準95%では、従うものと従わないものが存在したが、各データの平均を算出し、カイ2乗値を求めると、表2にあるように、カイ2乗値が $6.49 \times 10^{-7}$ となり、有意水準95%において、Benford's Lawに従う結果となった。

## 4 フィボナッチ数列

この章では、まず最初にフィボナッチ数列についての説明を行う。そして、Benford's Lawの割合と実測値に関してカイ2乗検定の結果を示す。最後に、その結果から得られたデータが、Benford's Lawに従うか否かを考察する。

### 4.1 定義

フィボナッチ数列は、数式(2)~(4)により定義される数式である。

$$F_1 = 1 \quad (2)$$

$$F_2 = 1 \quad (3)$$

$$F_{n+2} = F_n + F_{n+1} (n \geq 1) \quad (4)$$

この数式を用い計算していくと、

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377...

のようになり、 $n+2$ 項目の数値は、 $n$ 項目と $n+1$ 項目を足した数値となる。

### 4.2 結果

上記のフィボナッチ数列を求める数式を、 $n=1000$ としてプログラミングし、得られた結果とBenford's Lawの割合に関して、カイ2乗検定を行った。

表3 カイ2乗検定

	カイ2乗値
$n=100$	$6.88 \times 10^{-6}$
$n=500$	$6.13 \times 10^{-17}$
$n=1000$	$2.46 \times 10^{-19}$

### 4.3 考察

フィボナッチ数列は、人為的な意図が加わっていないかつ、完全にランダムな数列であるために、初見でBenford's Lawに当てはまると考えられた。ここでは、フィボナッチ数列 $n=100, 500, 1000$ の場合において、有意水準を95%と設定し、仮説検定を行う。

帰無仮説として、『フィボナッチ数列は、Benford's Lawに従わない』と仮定する。 $n=100, 500, 1000$ の時のカイ2乗値は、それぞれ、 $6.88 \times 10^{-6}, 6.13 \times 10^{-17}, 2.46 \times 10^{-19}$ である。

カイ2乗の95%点は、 $3.9 \times 10^{-3}$ であり、 $n=100, 500, 1000$ のいずれの場合も、カイ2乗値はカイ2乗の95%点より小さい結果となった。ゆえに、各 $n$ において、有意水準95%より大きい確率となる。

よって、帰無仮説が棄却され、対立仮説である『フィボナッチ数列は、Benford's Lawに従う』が採択される。

よって、フィボナッチ数列がBenford's Lawに従うと言えることが分かった。また、 $n$ の値が増加するにつれ、カイ2乗値が減少していくことから、項数が増加するにつれて、Benford's Lawの割合により近づくことが考えられる。

## 5 トリボナッチ数列

この章では、まず最初にトリボナッチ数列についての説明を行う。そして、Benford's Lawの割合と実測値に関してカイ2乗検定の結果を示す。最後に、その結果から得られたデータが、Benford's Lawに従うか否かを考察していく。

### 5.1 定義

トリボナッチ数列は、数式(5)~(8)により定義される数式である。

$$T_1 = 1 \quad (5)$$

$$T_2 = 1 \quad (6)$$

$$T_3 = 2 \quad (7)$$

$$T_{n+3} = T_n + T_{n+1} + T_{n+2} (n \geq 1) \quad (8)$$

この数式を用い計算していくと、

1, 1, 2, 4, 7, 13, 24, 44, 81, 149, 274, 504, 927, 1705...  
 のようになり,  $n + 3$  項目の数値は,  $n$  項目,  $n + 1$  項目,  
 $n + 2$  項目を足した数値となる.

## 5.2 結果

上記のトリボナッチ数列を求める数式を, フィボナッチ  
 数列と同様に,  $n=1000$  としてプログラミングした結果,  
 それぞれの 1 桁目の割合は表 4 の通りになる.

表 4 カイ 2 乗検定

	カイ 2 乗値
$n=100$	$2.01 * 10^{-4}$
$n=500$	$6.81 * 10^{-19}$
$n=1000$	$3.57 * 10^{-20}$

## 5.3 考察

トリボナッチ数列についても, 人為的な意図が加わって  
 いないかつ, 完全にランダムな数列であるために, 初見で  
 Benford's Law に当てはまると考えられた. ここでは, トリ  
 ボナッチ数列  $n=100, 500, 1000$  の場合において, 有意  
 水準を 95 % に設定し, 仮説検定を行う.

帰無仮説として, 『トリボナッチ数列は, Benford's Law  
 に従わない』と仮定する.  $n=100, 500, 1000$  の時のカイ 2  
 乗値は, 表 5 の通りで, それぞれ,  $2.01 * 10^{-4}, 6.81 * 10^{-19},$   
 $3.57 * 10^{-20}$  である.

カイ 2 乗の 95 % 点は,  $3.9 * 10^{-3}$  であり,  $n=100, 500,$   
 $1000$  のいずれの場合も, カイ 2 乗値は少ない結果となっ  
 た. ゆえに, 各  $n$  において, 有意水準 95 % より大きい確率  
 となる.

よって, 帰無仮説が棄却され, 対立仮説である 『トリボ  
 ナッチ数列は, Benford's Law に従う』が採択される.

以上のことから, トリボナッチ数列が Benford's Law に  
 従うことが証明できた. また,  $n$  が増加するにつれ, カイ 2  
 乗値が小さくなっていることから, 項数が増加するにつれ  
 て, Benford's Law の割合に近づくことが考えられる.

## 6 2 桁の Benford's Law

これまでは, 最初の 1 桁目の数値, つまり, 1 から 9 まで  
 の割合を考えてきた. ここでは, 最初の 2 桁の数値の 10 ~  
 99 までの数値の割合を考える. 1 桁の Benford's Law は  
 以前に述べた通り, 右肩下がりとなったが, その続きであ  
 る数値の割合はどのようになっているのかを述べていく.

次に 2 桁の Benford's Law の割合の求め方の定義を  
 示す.

### 6.1 定義

2 桁の Benford's Law は, (9) 式により求まる.

$$f(x, y) = \log_{10}\left(1 + \frac{1}{10x + y}\right) \quad (9)$$

ただし,  $x$  は, 例えば, 数値が 35 であれば 3 の 1 桁目の  
 数値を,  $y$  は, 5 の 2 桁目の数値を代入し求めることが可

能である. 1 桁目  $x=1,2,3...9$ , 2 桁目  $y=0,1,2...9$  である.  
 (9) 式により求まる 2 桁目の Benford's Law の割合も図 2  
 のように右肩下がり割合となる.

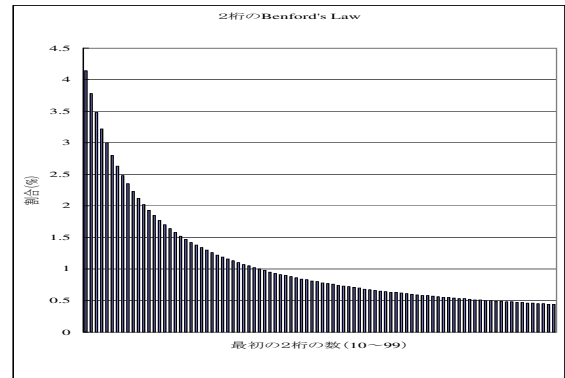


図 2 2 桁の Benford's Law

## 7 2 桁のフィボナッチ数列

フィボナッチ数列の最初の 1 桁の数値 1 ~ 9 の割合は,  $n$   
 を大きくするにつれて Benford's Law に従う結果となる  
 ことがわかった. 続いて, フィボナッチ数列の最初の 2 桁  
 の数値 10 ~ 99 の割合は Benford's Law に従うのか検証  
 した. 2 桁であるため標本数が多数必要なので  $n=1000$  で  
 本論文では検証する.

### 7.1 結果

フィボナッチ数列の最初の 2 桁の割合を求めた結果を  
 図 3 に示す. フィボナッチ数列の最初の 2 桁の割合が, (9)  
 式により求めた割合に従う結果となるか否か, カイ 2 乗検  
 定を行った結果が表 5 である.

表 5 カイ 2 乗検定

	カイ 2 乗値
$n=1000$	$1.57 * 10^{-18}$

### 7.2 考察

フィボナッチ数列  $n=1000$  の場合において, 有意水準を  
 95 % に設定し, 仮説検定を行う.

帰無仮説として, 『2 桁のフィボナッチ数列は, 2 桁の  
 Benford's Law に従わない』と仮定する.  $n=1000$  の時の  
 カイ 2 乗値は, 表 5 の通りで,  $2.46 * 10^{-19}$  である.

カイ 2 乗の 95 % 点は,  $3.9 * 10^{-3}$  であり,  $n=1000$  の場  
 合で, カイ 2 乗値は少ない結果となった. ゆえに, 各  $n$  に  
 おいて, 有意水準 95 % より大きい確率となる.

よって, 帰無仮説が棄却され, 対立仮説である 『2 桁の  
 フィボナッチ数列は, 2 桁の Benford's Law に従う』が採  
 択される.

以上のことから、最初の2桁のフィボナッチ数列の割合が2桁の Benford's Law に従うことが証明できた。

## 8 2桁のトリボナッチ数列

前章と同様に、ここでは  $n=1000$  における2桁のトリボナッチ数列について考える。

### 8.1 結果

トリボナッチ数列の最初の2桁の割合を求めた結果を図4に示す。トリボナッチ数列の最初の2桁の割合が、(9)式により求めた割合に従う結果となるか否か、カイ2乗検定を行った結果が表6である。

表6 カイ2乗検定

	カイ2乗値
$n=1000$	$1.57 * 10^{-18}$

### 8.2 考察

トリボナッチ数列  $n=1000$  の場合において、有意水準を95%に設定し、仮説検定を行う。

帰無仮説として、『2桁のトリボナッチ数列は、2桁の Benford's Law に従わない』と仮定する。 $n=1000$  の時のカイ2乗値は、表6の通りで、 $1.57 * 10^{-18}$  である。

カイ2乗の95%点は、 $3.9 * 10^{-3}$  であり、 $n=1000$  の場合で、カイ2乗値は少ない結果となった。

ゆえに、各  $n$  において、有意水準95%より大きい確率となる。

よって、帰無仮説が棄却され、対立仮説である『2桁のトリボナッチ数列は、2桁の Benford's Law に従う』が採択される。

以上のことから、トリボナッチ数列が Benford's Law に従うことが証明できた。

## 9 $n$ 進数の Benford's Law

これまででは、10進数における Benford's Law を考察してきたが、ここでは視点を変え、 $n$ 進数 ( $n = 2, 3, \dots, 9$ ) の時に、Benford's Law を適応した場合、どのような割合になるかを論ずる。また10進数で成立した Benford's Law が、 $n$ 進数に変化しても、成立するかを研究した。各  $n$ 進数に対し、(1)式によって求まる最初の桁の数値の割合を表7、図3に示す。 $n$ 進数に変換した場合に関して、右肩下がり割合になる。

表7  $n$ 進数の割合 (単位:%)

	1	2	3	4	5	6	7	8
2進数	100							
3進数	63.1	36.9						
4進数	50.0	29.2	20.8					
5進数	43.1	25.2	17.9	13.9				
6進数	38.7	22.6	16.1	12.5	10.2			
7進数	35.6	20.8	14.8	11.5	9.4	7.9		
8進数	33.3	19.5	13.8	10.7	8.8	7.4	6.4	
9進数	31.5	18.5	13.1	10.2	8.3	7.0	6.1	5.4

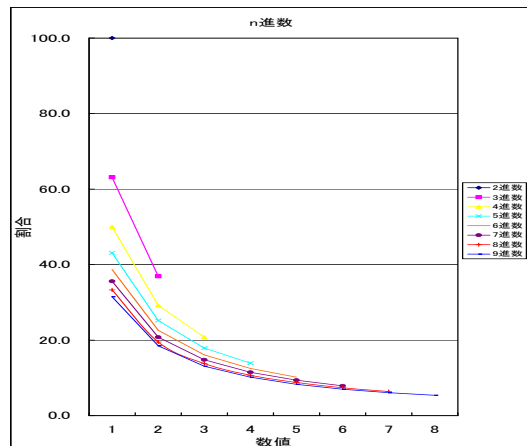


図3  $n$ 進数の割合

### 9.1 結果

$n$ 進数の Benford's Law の検証で用いるものとして、10進数で考えた場合、Benford's Law に最も従う結果となったフィボナッチ数列を用いる。フィボナッチ数列を  $n$ 進数に変換し、その最初の桁の数値の割合が、表7の割合に従うか否かを、カイ2乗検定を用い考察する。その結果を表8、表9に示す。

表8 カイ2乗検定 (項数:100)

	カイ2乗値
3進数	$3.71 * 10^{-6}$
4進数	$9.09 * 10^{-7}$
5進数	$1.5 * 410^{-7}$
6進数	$1.66 * 10^{-13}$
7進数	$4.27 * 10^{-14}$
8進数	$3.32 * 10^{-16}$
9進数	$8.19 * 10^{-16}$

表9 カイ2乗検定 (項数:500)

	カイ2乗値
3進数	$1.04 * 10^{-4}$
4進数	$1.90 * 10^{-9}$
5進数	$7.38 * 10^{-12}$
6進数	$8.64 * 10^{-16}$
7進数	$1.23 * 10^{-16}$
8進数	$2.66 * 10^{-25}$
9進数	$1.16 * 10^{-28}$

### 9.2 考察

ここで、有意水準を95%に設定し、仮説検定を行う。帰無仮説として、『 $n$ 進数の Benford's Law の割合に、 $n$ 進数のフィボナッチ数列における Benford's Law の割合は

従わない』と仮定する。項数 100, 500 における、各進数のカイ 2 乗値の最大値でも  $1.0 \times 10^{-4}$  であり、これは 95 % 点の  $3.9 \times 10^{-3}$  より小さい数値である。ゆえに、すべての進数において有意水準 95 % より大きい確率となる。有意水準を 95 % に設定したために、これは帰無仮説が棄却され対立仮説である『 $n$  進数の Benford's Law の割合に  $n$  進数のフィボナッチ数列における Benford's Law の割合は従う』が採択される。以上のことから、 $\log$  計算より算出した  $n$  進数の Benford's Law に従う結果となった。項数 100 の時より 500 の時の方が、カイ 2 乗値が小さいため、項数が増大すればする程 Benford's Law に限りなく近づくことがわかった。

2 進数に関しては、最初の桁の数値が 1 にしかならないために、数式で求めた 2 進数の割合と合致する。

## 10 Benford's Law に従わないデータ

この章では、素数データ、乱数データの説明をして、そのプログラミングの結果となぜ素数、乱数が Benford's Law に従わないかを考察していく。

### 10.1 素数のデータ

Benford's Law に従わないデータとして、素数を扱う。一見素数も Benford's Law に従う乱数であると考えられる。しかし、プログラミングを組み検証してみたところ、従わない結果となった。

### 10.2 結果 1

素数の一桁目の割合をプログラミングで算出した結果、図 4、図 5 となる。

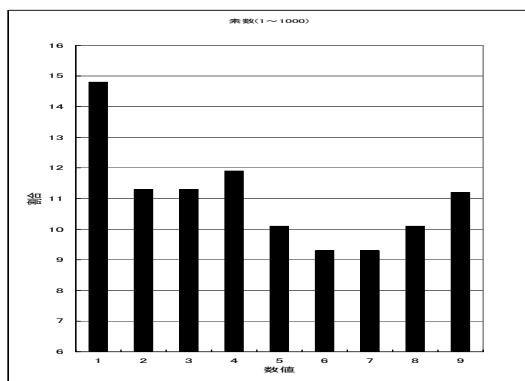


図 4 素数の 1 桁目の割合 (1~1000)

### 10.3 結果 2

素数の割合は、フラットな割合となると考察していた。そのことから、次に素数データの最初の桁の数値の割合が、完全に等分布となる一様分布になると仮定し、素数データの実測値と一様分布についてカイ 2 乗検定を行った。その検定結果を表 10 に示す。

### 10.4 考察

初見では、素数の最初の桁の数値は、ランダムな数列で Benford's Law に従うと考えられた。しかし、実際にプログラミングを組み、解析してみたところ、結果の図を見て

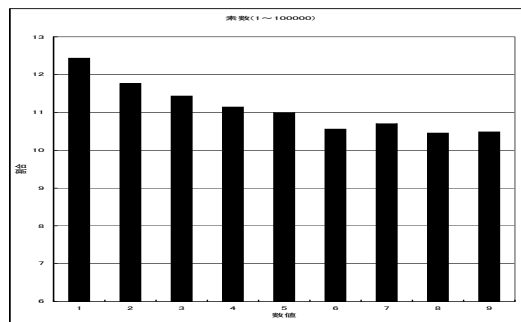


図 5 素数の 1 桁目の割合 (1~100000)

表 10 カイ 2 乗検定

	カイ 2 乗値
素数 (1~1000)	$1.06 \times 10^{-2}$
素数 (1~10000)	$2.58 \times 10^{-4}$
素数 (1~100000)	$7.30 \times 10^{-5}$

もわかるように、Benford's Law に従わない結果となった。

その原因として考えられるものを次に述べる。素数を算出するアルゴリズムは、2 の倍数、3 の倍数と素数の倍数を切り捨て、残ったものが素数となる。数値が大きくなっていくにつれて、割り切れる数が増えていく。

1~1000 で考えた場合、100 以上になると割り切れる数はほぼ均一になる。100~199, 200~299, 300 から... とどの 1 桁目も数値の幅は 100 であるため、差はあまりないと考えられる。1~100000 で考えた場合、1~1000 で考えた時より割合がフラットになることから、そのことは明白である。

扱う数値を増やせば増やすほど素数の最初の桁数は 1~9 まで横ばい状態に近づくと考えられる。

素数における最初の桁の数値の割合は、中間発表では、視覚的のみで横ばいの割合となっていると述べた。しかし、実際に素数の割合と一様分布の割合についてカイ 2 乗検定を行ったところ、1~1000 の素数に関するカイ 2 乗値は  $1.06 \times 10^{-2}$  と予想を反して、一様分布に従わない結果が得られた。

素数データは、有意水準 95 % においては、帰無仮説である、『素数 1~1000 の最初の桁の割合は、一様分布ではない』が採択される。

次に、素数のデータを 1~10000 まで増やし、カイ 2 乗検定を行ったところ、カイ 2 乗値が  $2.58 \times 10^{-4}$  となり、1~1000 の時の素数のデータよりも、より一様分布に近くなると結果が得られた。

また、1~100000 の素数のカイ 2 乗値が、 $7.30 \times 10^{-5}$  であることから、素数の項数が増えるごとに、最初の桁の数値は、より等分布に近づくことが数学的根拠に基づきわかる。

## 10.5 乱数データ

扱うデータとして人為的な意図が加わっていない乱数を発生させ、その1桁目の数値の割合が、どのような割合となるか述べる。初見では、乱数を発生させるために、それぞれの数値の割合が均等になることが考えられた。次に、その結果を示す。

## 10.6 結果 1

乱数を発生させるものとして Excel の RANDBETWEEN というコマンドを用いた。その数値の割合を出した際の1桁の割合を図6に示す。

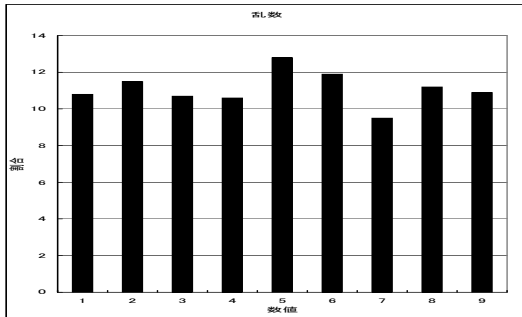


図6 乱数の割合

## 10.7 結果 2

次に乱数データの最初の桁の数値の割合が、完全に等分布となる一様分布になると仮定し、乱数データの実測値と一様分布に関してカイ2乗検定を行った。その結果が表11となる。

表11 カイ2乗検定

	カイ2乗値
乱数	$8.66 * 10^{-4}$

## 10.8 考察 1

Benford's Law とは、人為的な意図が加わっていないデータを用いた場合、ある一定の割合に限りなく近づくとする法則であった。

一様乱数を発生させる Excel のコマンド RANDBETWEEN を使い、乱数の1桁目の数値の割合が Benford's Law に従うか否か検定を行った。図6からもわかる通り、人為的な意図が加わっていないデータでも Benford's Law に従わない結果となった。また、この横ばいの図は、素数データの図4、図5に近いものと言える。従わない結果として考えられる点は、いくら人為的な意図が加わっていないとは言え、完全にランダムな乱数をとった場合は、各々の数の割合が均等になることである。

## 10.9 考察 2

考察2では、乱数データが一様に分布しているか否かを、数学的根拠に基づき論ずる。カイ2乗値が  $8.66 * 10^{-4}$

となり、有意水準 95 %において、帰無仮説を、『発生させた乱数が一様分布に従わない』とすると、帰無仮説が棄却され、対立仮説である、『発生させた乱数が一様分布に従う』が採択される。

このことから、乱数データは一様分布に従い、横ばいであると数学的根拠に基づき言える。

## 11 まとめ

Benford's Law とは、人為的な意図が加わっていないデータにおける、最初の桁の数値の割合は、ある一定の割合に近似するという法則であった。フィボナッチ数列等の数列式のデータが Benford's Law の割合に最も近づく結果が得られた。このことから、最初の桁の数値の割合が Benford's Law に近似する原因として考えられることは、桁数が繰り上がる際に1~4の数値は、出現する頻度が高いためであると言えるのではなからうか。

また物理学者ベンフォードが、『桁数が繰り上がる実際の統計データを用いる』という人為的な意図を加えた結果、ある一定の割合に近づいたと論じた法則であると考えられる。このことは、人為的な意図が加わっていない乱数データが、Benford's Law に従わないことから明白である。

このように Benford's Law とは、私達の全く意識しないところで、息づいている法則であると言え、今回、本論文の Benford's Law を研究することにより、今まで見ていなかった統計の世界が見えた。

## 12 参考文献

- [1] Benford's Law and the 2004 Presidential Election, <http://gabriel.physics.ucsb.edu/pagem/Benford/>
- [2] WIKIVERSE Benford's Law, <http://benford-s-law.wikiverse.org/>
- [3] 白旗慎吾, 「統計解析入門」, 共立出版 (1992,5)
- [4] 経済関連ランキング, <http://www.hyou.net/menu/keizai.htm/>
- [5] 田中英雄, 「日本都市年鑑 平成 17 年度版」, 全国市長会 (2005,12)
- [6] 総務省統計局, 「世界の統計 2006 年版」, 総務省統計研修所 (2006,3)