

「正規数に関する統計的数値実験と自然言語に現れる数の偏り」

理学部数学科

秋山 正和

1 4 7 1 0 0 3 J

## 目次

- 0.1 … 正規数とは？
- 0.2 … 数値実験
- 0.3 … 計算結果
- 0.4 … データの解析
- 0.5 …  $\chi^2$  検定での解析
- 0.6 … 信頼区間
- 0.7 … 自然言語に現れる数の分布
- 0.8 … 反省

### 0.1 正規数とは？

$x = \sum x_k r^{-k}$  を、実数  $x$  の  $r$  進展開とする。  $0, 1, \dots, r-1$  の任意の  $n$  個組  $B_n = (b_1, b_2, \dots, b_n)$  に対して、  $N_k(x, b_n)$  を  $x_1, \dots, x_k$  のなかに現れる  $B_n$  の度数とする。すべての  $n, B_n$  に対して  $N_k(x, b_n)/k \rightarrow r^{-n}$  ( $k \rightarrow \infty$ ) のとき、  $x$  を  $r$  進正規数と呼ぶ。殆ど全ての実数は任意の  $r$  に対して正規数である。例えば 2 進正規数の具体例は  $0.110111001011101111000\dots$  により構成できる。ところが  $\sqrt{2}$ 、 $\pi$ 、 $e$  などの具体的な無理数が正規数かどうかは不明である。

### 0.2 数値実験

以下の C 言語プログラムによりファイルから配列  $A$  に小数点第一桁目小数点第二桁目、 $\dots$  が入り、配列  $A$  に値が入るごとに配列  $B_1$  に 0 から 9 までの個数が、  $B_2$  に 0 から 99 までの個数が、  $B_3$  に 0 から 999 までの個数が、  $B_4$  に 0 から 9999 までの個数がある。このプログラム上では 100 万 - 4 桁まで調べている。調べた数は  $\sqrt{2}$ 、 $e$ 、オイラー定数、 $\pi$ 、 $\log 2$  である。

例  $\sqrt{2}$  を例にとると、 $B_1[0] = (100 \text{ 万} - 4 \text{ 桁までの } 0 \text{ の個数}) = 99813$ .  
プログラムは

```
#include <stdio.h>
#include <stdlib.h>
#define N 13
#define M 10000
int myatoi(char s)
{
    switch(s){
        case '0':
            return 0;
            break;
        case '1':
            return 1;
            break;
        case '2':
            return 2;
            break;
        case '3':
            return 3;
            break;
        case '4':
            return 4;
            break;
        case '5':
            return 5;
            break;
        case '6':
            return 6;
            break;
        case '7':
            return 7;
            break;
        case '8':
            return 8;
            break;
        case '9':
            return 9;
            break;
        default:
            return 11;
    }
}
```

```

}

main()
{
    FILE *fin;
    char filename1[20];
    char c;
    int i=0,A[N],j,k=0,l=1,d;
    int B1[10], B2[100], B3[1000], B4[10000];
    printf("file name?");
    gets(filename1);
    fin=fopen(filename1,"r");
    if(fin==NULL){
printf("can not open %s\n",filename1);
exit(1);
    }
    for(k = 0;k < 10;k++){
B1[k] = 0;
    }
    for(k = 0;k < 100;k++){
B2[k] = 0;
    }
    for(k = 0;k < 1000;k++){
B3[k] = 0;
    }
    for(k = 0;k < 10000;k++){
B4[k] = 0;
    }
while(1){
    for(i = 1; i < N; i++){
        d = -1;
        while((c = fgetc(fin)) != EOF){
            d = myatoi(c);
            if(d != 11){
                break;
            }
        }
        if(d != -1){
            A[i] = d;
            d = -1;
        }
        else{
            /* EOF */

```

```

        break;
    }
}
if(i == N){
    for(j = 1; j < N; j++){
        B1[A[j]]++;
    }
    for(j = 1; j < N; j+=2){
        B2[A[j]*10 + A[j+1]]++;
    }
    for(j = 1; j < N; j+=3){
        B3[A[j]*100 + A[j+1]*10 + A[j+2]]++;
    }
    for(j = 1; j < N; j+=4){
        B4[A[j]*1000 + A[j+1]*100 + A[j+2]*10 + A[j+3]]++;
    }
}
else{
    break;
}
}

fclose(fin);
}

```

### 0.3 計算結果

データが膨大なため、それぞれの数のにおいて、一桁区切りと二桁区切りのみ表示する。表は

0 | 1 | ... | 9

00 | 01 | 02 | ... | 09

01 | ...

の順で並んでいる。

root2

99813 | 98924 | 100435 | 100190 | 100023 | 100155 | 99886 | 100008 | 100441 | 100121 |

total = 999996

5001 | 5050 | 4927 | 4973 | 5157 | 5003 | 4882 | 5102 | 5025 | 4963 | 4861 | 4931 | 4968 |

4940 | 4998 | 4856 | 4855 | 4850 | 4971 | 4987 | 4966 | 4999 | 5101 | 5102 | 5013 | 4991 |

4919 | 4996 | 5011 | 5155 | 5028 | 5067 | 4925 | 5064 | 5025 | 4953 | 4962 | 5006 | 5148 |

4944 | 5026 | 4863 | 4913 | 5043 | 4923 | 5047 | 4970 | 5016 | 5051 | 5010 | 4966 | 4963 |

5152 | 5122 | 5041 | 4858 | 5025 | 4971 | 5095 | 5121 | 5036 | 4954 | 5094 | 4978 | 4909 |

5093|5101|5003|5101|5111|4839|5060|5024|4897|5121|5036|4908|5073||  
4943|5080|5042|4993|4982|5052|4904|5071|4947|5114|5011|4994|4965||  
4827|5096|4897|5070|4933|4937|4896|4975|5080|total = 499998

e

99425|100131|99845|100228|100389|100087|100479|99910|99811|99691||  
total = 999996  
4951|4937|4898|4896|5114|4945|5001|5013|4974|4944|4936|5098|5026||  
4936|5114|4859|5053|5101|4988|4911|4940|4914|4997|5121|4971|5039||  
5157|4903|5034|5004|4868|5027|4976|5084|4847|5085|5016|4963|4948||  
5065|4995|5095|4947|4974|5013|5028|5117|5002|4989|5123|5013|5105||  
5049|5165|4949|4904|5008|4963|5024|5006|5092|4962|4961|5048|5035||  
5015|5011|4945|5039|4988|4973|4940|5045|5036|5047|5000|5043|4974||  
5078|5060|4876|4938|4987|5049|4998|4986|5065|4979|4965|4831|5108||  
5093|4879|5040|5018|5040|4912|4871|5098|4850|total = 499998

Euler

100150|100143|99794|100194|99783|100110|100169|99681|100135|99837||  
total = 999996  
5109|5134|5079|4990|4931|4970|5080|4896|4912|5029|4955|4970|4998||  
5072|4995|4921|5091|5028|5084|5068|4860|4971|5027|5038|5054|5037||  
5063|4845|5010|4944|5042|5103|4877|5065|4951|5099|4958|5009|4949||  
5117|5027|4932|5053|5016|4868|4970|5029|4946|5048|4863|4988|5004||  
4979|4899|5027|4970|5041|5049|5023|5022|4898|4807|4986|4974|5067||  
5114|5055|4945|4974|4919|5020|5042|5041|5048|5012|5110|5001|4974||  
5047|4894|5060|5061|5040|4999|5021|4991|5005|4810|5047|4945|5061||  
4937|4865|4923|5105|4926|5107|4990|5062|5030|total = 499998

Pi

99959|99756|100026|100229|100230|100358|99548|99800|99984|100106||  
total = 999996  
4970|4974|4999|5102|5000|4969|4892|4959|5103|5082|4953|5055|5010||  
5104|5002|4940|4885|5041|5121|5086|4864|5021|5018|4901|4991|5050||  
5018|5148|4996|4924|5069|4977|5087|5059|4868|5056|4910|5061|4968||  
5082|4917|4968|4968|4953|4976|5039|5042|5021|4977|5057|5033|4888||  
5024|5033|5092|5120|5026|5063|4900|5023|5121|4876|5051|4954|4962||  
5034|4885|5039|4953|4917|4984|4871|4970|4985|5165|4892|5113|4887||  
4934|4847|5043|4978|4958|5010|5049|5019|5052|4967|4988|5026|4955||  
4951|5010|4991|5207|5037|4933|4966|4954|5029|total = 499998

log2

100154|100300|99496|99995|99734|100221|99863|99857|99832|100544||  
total = 999996  
5045|5133|4972|5082|4880|5048|5055|4897|4996|5088|4990|5040|4967||  
4911|5071|5059|4963|4870|5093|5095|4966|4973|5002|4850|5137|4902||  
4891|4862|5084|4975|4989|5067|4973|4944|4990|4994|4985|5028|5028||  
5025|5079|5068|4872|5049|4924|4965|4955|4972|4961|5015|5027|5009||

5004|4966|4982|4986|5082|5119|4964|5094|4980|4963|4921|4953|5134|  
 4967|5110|4974|4959|5067|4940|5081|5101|5009|4880|5097|4920|5052|  
 4971|4981|4924|4980|5033|5094|4912|4967|4943|4952|4921|5084|5018|  
 4927|5009|5114|4964|5003|4931|5099|5045|5005|total = 499998

## 0.4 データの解析

期待されたように、どの数も一桁、二桁、三桁、四桁に区切ると、10%、1%、0.1%、0.01%ぐらいの頻度で現れるが、区切る数を増やすごとに誤差も大きくなっていることが分かった。 $\sqrt{2}$ の一桁と二桁のみの頻度(%)を下に示す。

9.981340|9.892440|10.043540|10.019040|10.002340|10.015540|9.988640|10.000840|10.044140|10.012140|  
 1.000204|1.010004|0.985404|0.994604|1.031404|1.000604|0.976404|1.020404|  
 1.005004|0.992604|0.972204|0.986204|0.993604|0.988004|0.999604|0.971204|  
 0.971004|0.970004|0.994204|0.997404|0.993204|0.999804|1.020204|1.020404|  
 1.002604|0.998204|0.983804|0.999204|1.002204|1.031004|1.005604|1.013404|  
 0.985004|1.012804|1.005004|0.990604|0.992404|1.001204|1.029604|0.988804|  
 1.005204|0.972604|0.982604|1.008604|0.984604|1.009404|0.994004|1.003204|  
 1.010204|1.002004|0.993204|0.992604|1.030404|1.024404|1.008204|0.971604|  
 1.005004|0.994204|1.019004|1.024204|1.007204|0.990804|1.018804|0.995604|  
 0.981804|1.018604|1.020204|1.000604|1.020204|1.022204|0.967804|1.012004|  
 1.004804|0.979404|1.024204|1.007204|0.981604|1.014604|0.988604|1.016004|  
 1.008404|0.998604|0.996404|1.010404|0.980804|1.014204|0.989404|1.022804|  
 1.002204|0.998804|0.993004|0.965404|1.019204|0.979404|1.014004|0.986604|  
 0.987404|0.979204|0.995004|1.016004|

しかしこの解析の仕方ではこれ以上のことは分からなかった。

## 0.5 $\chi^2$ 検定での解析

$\sqrt{2}$  を例にとる。今  $\sqrt{2} = 1.4142135 \dots$  の 4,1,4,2,1,3... を独立で一様な確率変数列と仮定する。

$\chi^2 = \sum \frac{\text{自由度} \cdot (\text{実測値} - \text{理論値})^2}{\text{理論値}}$  を計算する。

$$\chi_1^2 = \sum_{k=0}^9 \frac{(B_1[k] - \frac{999996}{10})^2}{\frac{999996}{10}}$$

$$\chi_2^2 = \sum_{k=0}^{99} \frac{(B_2[k] - \frac{499998}{100})^2}{\frac{499998}{100}}$$

$$\chi_3^2 = \sum_{k=0}^{999} \frac{(B_3[k] - \frac{333332}{1000})^2}{\frac{333332}{1000}}$$

$$\chi_4^2 = \sum_{k=0}^{9999} \frac{(B_4[k] - \frac{249999}{10000})^2}{\frac{249999}{10000}}$$

\*添え字は区切りの桁数を表す。 計算結果は

表 1: それぞれの数の  $\chi^2$  の値

/	$\sqrt{2}$	e	Euler		Log2
$\chi_1^2$	16.64	9.51	3.57	5.51	8.50
$\chi_2^2$	127.73	107.76	103.86	105.88	96.16
$\chi_3^2$	1056.59	976.85	1051.21	998.11	1028.19
$\chi_4^2$	10056.66	9932.11	10216.04	987162	10073.90

確率変数 X が自由度 n の  $\chi^2$  分布に従う  $\Leftrightarrow$

X の確率密度関数が

$$f(x) = \begin{cases} \frac{e^{-x/2} x^{n/2-1}}{\Gamma(\frac{n}{2}) 2^{n/2}} & (x > 0) \\ 0 & (x \leq 0) \end{cases} \quad (1)$$

で与えられる。

これを元に、 $100 \times (1 - \int_0^{\text{表の値}} f(x) dx)$  を計算する。

表 2: それぞれの数の  $\chi^2$  の値の p 値

/	$\sqrt{2}$	e	Euler		Log2
$\chi_1^2$	5.5	48.4	93.7	78.8	48.4
$\chi_2^2$	2.8	25.7	35.0	30.0	56.2
$\chi_3^2$	10.0	8.6	12.2	50.2	25.4
$\chi_4^2$	34.0	68.1	6.3	81.6	30.0

値は%で示している。

この値が、0パーセントに近い場合は数の分布が理想から離れていると思える。また逆に100パーセントに近い場合は数の分布が理想に近すぎると推察できる。今、表から $\sqrt{2}$ の $\chi_2^2$ のp値は2.8である。このことについて考察する。200万桁-8桁まで $\sqrt{2}$ の近似値を取るとき、この値は小さいままであろうか? 数値実験の結果以下のように00から99までの数がえられた。

10032|| 9930|| 9875|| 9997||10208||10051||10009||10014|| 9972||10002  
 9912|| 9922|| 9950||10023|| 9944|| 9823|| 9975|| 9980|| 9936|| 9860  
 9990|| 9995||10006|| 9985||10032|| 9854||10023||10013|| 9968||10160  
 9927||10022||10074||10072||10033||10063|| 9933|| 9942||10187|| 9907  
 10010|| 9827||10064||10081|| 9936||10154|| 9954|| 9983||10063|| 9950



9824|| 9878||10180|| 9992||10039|| 9923|| 9979|| 9908|| 9972||10159  
 10060|| 9921||10250|| 9928|| 9908||10222||10149||10153||10046||10052  
 9848||10082|| 9841|| 9823||10144||10119|| 9965||10025|| 9915||10206  
 10002||10026|| 9980|| 9957|| 9995||10078||10172||10060||10014||10062  
 9941|| 9755||10023|| 9852||10097|| 9934|| 9993|| 9930|| 9879||10112

この表の数について、同様の計算を行う。

$$\chi^2_2 = \sum_{k=0}^{99} \frac{(B_2[k] - \frac{999996}{100})^2}{\frac{999996}{100}} = 100.235$$

よって

$$100 \times (1 - \int_0^{100.235} f(x) dx) = 44.64$$

であることが分かる。

つまり 200 万桁まで調べたとき、これら 5 つの数が正規数ではないとは言い切れないのである。

## 0.6 信頼区間

$\sqrt{2}$  を例にとる。0 の出る確率は 10% ぐらいであることは分かるが、どれほどの信頼度があるものであろうか？ 今 0 の出る確率を  $p$ 、1 から 9 までが出る確率を  $1 - p$  とする。この試行を  $n$  回繰り返したとき、一回一回の試行が独立であると仮定するとき、0 がでる回数を表す確率変数を  $X$  とする。このとき  $X = k$  となる確率は

$P(X = k) = {}_n C_k p^k (1 - p)^{n - k}$  である。 $X$  はこれを確率密度関数にもつ 2 項分布となる。一般に  $n$  が十分大なるとき、(応用上は  $n \geq 60$ ) 2 項分布  $B(n, p)$  は正規分布  $N(np, \sqrt{np(1 - p)})$  に分布収束する。つまり  $X \sim N(np, \sqrt{np(1 - p)})$  ( $n; 大$ ) ここで 2 項分布  $B(n, p)$  の期待値、分散はそれぞれ

$$E(X) = np$$

$$V(X) = np(1 - p)$$

である。確率変数  $X$  を一次変換して

$$Y = \frac{X - E(X)}{\sqrt{V(X)}}$$

とすれば、 $Y \sim N(0, 1)$  である。

いま標準正規分布の 99.9% 点は 3.291、つまり  $P(|Y| \leq 3.291) = 0.999$  である。

ここで試行の回数  $n$  は 999996 であり、999996 回までに 0 の出た回数  $k$  はデータから  $k = 99813$  である。 $p$  の点推定として  $p = \frac{k}{n} \approx 0.09981$  を用いると、 $\sqrt{V(x)} \approx 299.7193$  より、99.9% の信頼度で  $k - 3.219\sqrt{V(x)} \leq E(x) \leq k + 3.219\sqrt{V(x)}$  である。

整理すると  $p$  の範囲は

$0.09904 \leq p \leq 0.100587$  となる。このようにするとある数字が出る確率の範囲を、任意の信頼度で調べることができる。

また、例えば 0 が出る確率を  $p$ 、1 が出る確率を  $q$ 、2 から 9 までが出る確率を  $1 - p - q$  とする。すると 0 が出る回数と 1 が出る回数を表す分布は 3 項分布となる。これもまた 2 次元の正規分布に  $n$  が十分大なるとき、分布収束することが認められるので、上のような議論を使えば 0 が出る確率の  $p$  と 1 が出る確率の  $q$  の範囲を (ただし 2 次元的な領域となる) 任意の信頼度で求めることができる。

このようなことを  $\sqrt{2}$ 、e、オイラー定数、 $\pi$ 、Log 2 それぞれの数に対して適用し、0 から 9 までの出る確率の任意信頼度の信頼区間、0 0 から 9 9 まで出る確率の任意信頼度の信頼区間、... を求めることは可能である。

## 0.7 自然言語に現れる数の分布

自然言語は人間が話す言葉や、新聞、広告の値段など満ち溢れている。人が使う言葉なのでそこには自然に偏りが生まれるであろう。特に経済などでよく使う数字は0や1であろう。この仮説を調べる。次は信頼できるサイト (yahoo,livedoor, および EXCITE) からダウンロードしたある3日間の経済のデータである。

yahoo

順位 コード 市場 名称 取引値 前日比 (パーセント) 出来高

1 9658 JASDAQ (株) ビジネスブレイン太田昭和 1/14 14:56 400 +75 +23.08 23,000  
 2 6709 東証 2 部 明星電気 (株) 1/14 15:00 258 +47 +22.27 2,850,000  
 3 5962 大証 2 部 浅香工業 (株) 1/14 14:57 235 +42 +21.76 86,000  
 4 5018 JASDAQ MORESCO 1/14 15:00 13,400 +2,000 +17.54 4,900  
 5 1765 名証 イビデングリーンテック (株) 1/14 13:26 281,000 +40,000 +16.60 5  
 6 4974 マザーズ タカラバイオ (株) 1/14 15:00 283,000 +40,000 +16.46 26,824  
 7 4311 JASDAQ (株) ディースリー・パブリッシャー 1/14 15:00 292,000 +40,000 +15.87 167  
 .  
 .  
 .

livedoor

順位, 市場, コード, 会社名, 取引値, 前日比 (パーセント), 出来高, 売買代金 (100 万円), 時価総額 (億円)

1 JASDAQ 9658 ビジネスブレイン太田昭和 400 75 23.08 23,000 9 32  
 2 東証 2 部 6709 明星電気 258 47 22.27 2,850,000 735 154  
 3 大証 2 部 5962 浅香工業 235 42 21.76 86,000 20 26  
 4 JASDAQ 5018 MORESCO 13,400 2,000 17.54 4,900 65 96  
 5 東証マザーズ 4974 タカラバイオ 283,000 40,000 16.46 26,824 7,591 741  
 6 JASDAQ 4311 ディースリー・パブリッシャー 292,000 40,000 15.87 167 48 40  
 .  
 .  
 .

EXCITE

順位 コード 銘柄名 値上がり率 (%) 取引値 (円)

1 8902/東 1 パシフィックマネジメント (株) +16.17 273,000  
 2 6375/東 1 日本コンベヤ (株) +12.33 164  
 3 8840/東 1 (株) 大京 +10.42 212  
 4 7211/東 1 三菱自動車工業 (株) +10.20 162  
 5 1503/東 1 住友石炭鉱業 (株) +9.66 159

6 4321/東1 ケネディ・ウィルソン・ジャパン (株) +9.60 194,000

7 6508/東1 (株) 明電舎 +8.94 256

・  
・  
・

この3つのデータの0から9までの個数と割合を調べた結果を下の表に示す。

表3: yahoo の場合の数の分布 (0 から 9 まで)

計 63888 個	0	1	2	3	4	5	6	7	8	9
個数	13195	13930	5763	4512	5812	5739	3469	4074	3987	3407
パーセント	20.65	21.80	9.02	7.06	9.10	8.98	5.43	6.38	6.24	5.33

表4: livedoor の場合の数の分布 (0 から 9 まで)

計 51252 個	0	1	2	3	4	5	6	7	8	9
個数	10647	7946	5259	4748	4232	4130	3844	3797	3573	3076
パーセント	20.77	15.50	10.26	9.26	8.26	8.06	7.50	7.41	6.97	6.00

表5: EXCITE の場合の数の分布 (0 から 9 まで)

計 15204 個	0	1	2	3	4	5	6	7	8	9
個数	2285	2061	2002	1644	1483	1214	1159	1179	1114	1063
パーセント	15.03	13.56	13.17	10.81	9.75	7.98	7.62	7.75	7.33	6.99

2桁ごとに区切った場合、つまり00～99までの個数表は次に示す。ただしi行j列( $0 \leq i \leq 9, 0 \leq j \leq 9$ )は $10*i+j$ の個数を表す。

これらをヒストグラムにしたものを次に示す。表は yahoo,livedoor,EXCITE の順である。

表 6: yahoo の場合の数の分布 (00 から 99 まで)

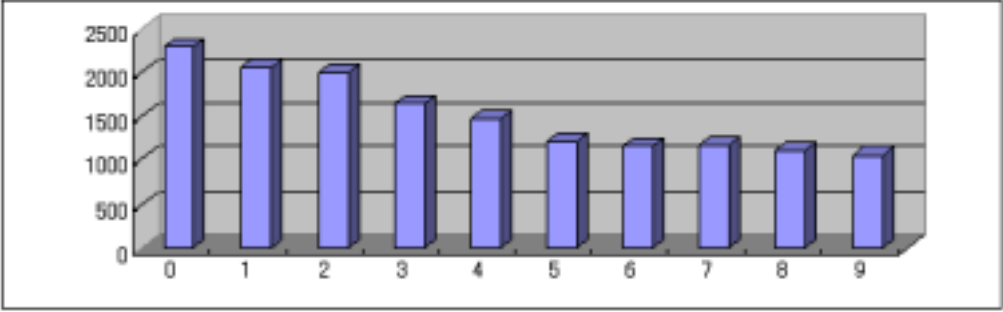
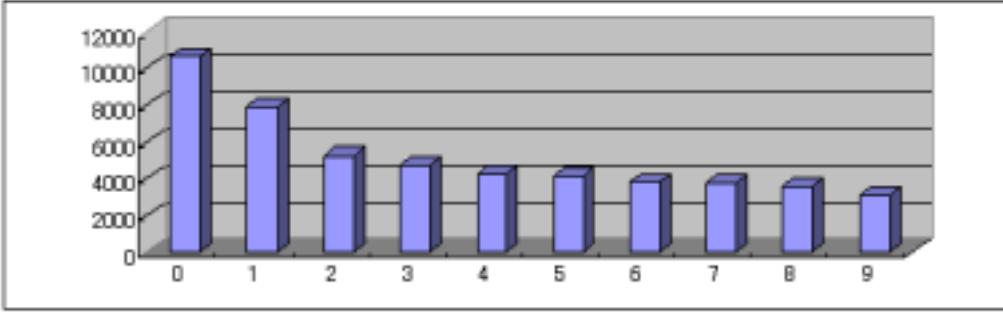
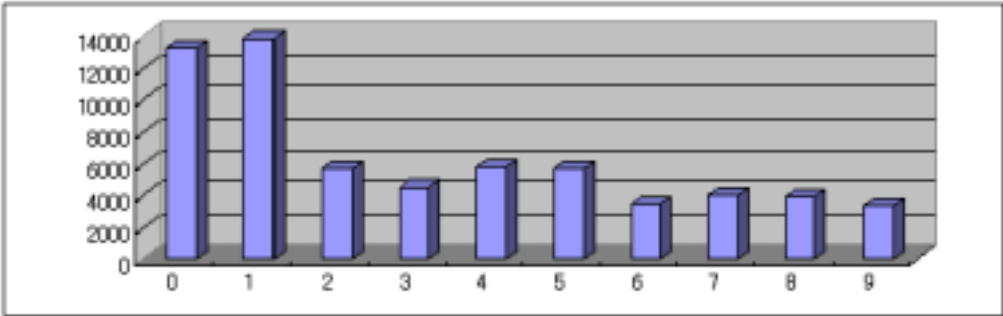
計 63888 個	0	1	2	3	4	5	6	7	8	9
0	3049	824	703	466	458	250	250	192	194	217
1	585	2034	354	293	1122	834	235	629	625	201
2	371	795	235	258	187	209	193	219	160	179
3	340	346	269	233	186	214	201	181	171	144
4	337	706	265	195	220	585	170	166	159	152
5	857	407	242	188	177	160	176	158	202	318
6	240	255	212	180	139	143	110	158	148	127
7	271	650	247	120	138	143	150	101	96	101
8	266	667	207	130	107	170	126	120	102	112
9	276	334	223	164	123	146	146	133	123	94

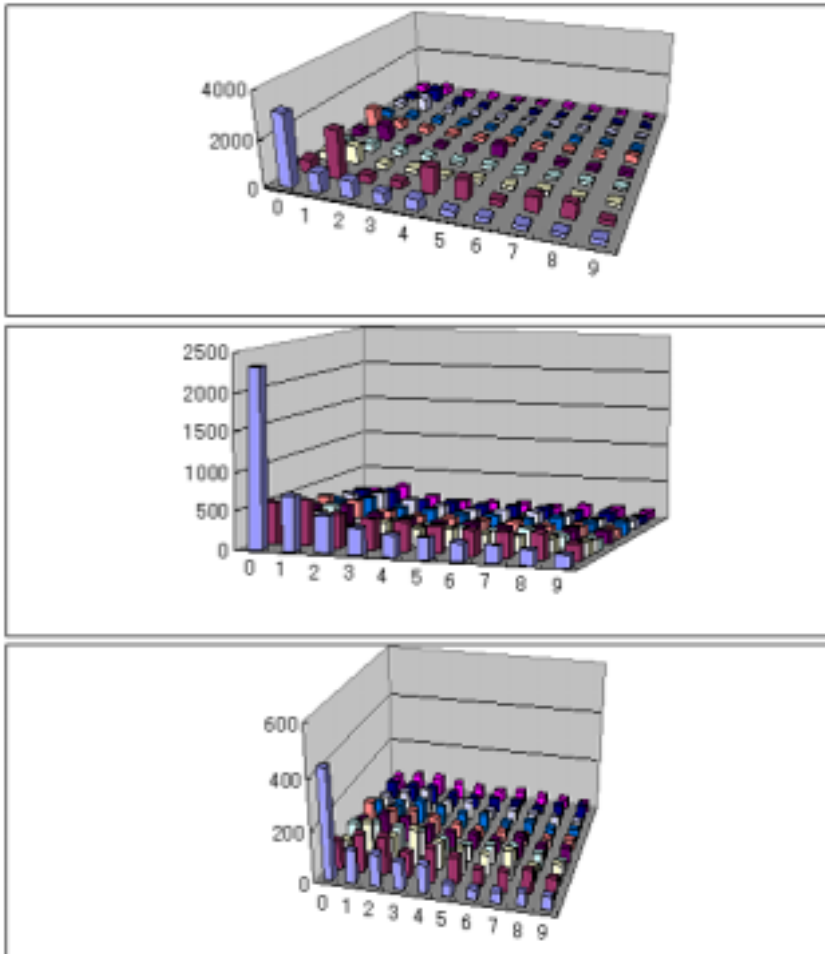
表 7: livedoor の場合の数の分布 (00 から 99 まで)

計 51252 個	0	1	2	3	4	5	6	7	8	9
0	2327	713	485	339	294	279	246	229	203	164
1	541	600	439	402	384	357	326	323	344	231
2	398	397	220	293	234	246	223	215	206	187
3	335	407	239	248	215	208	184	231	188	144
4	330	331	245	209	176	194	167	161	170	172
5	375	337	223	178	183	158	164	146	169	136
6	277	315	206	201	161	165	146	185	148	157
7	281	310	247	185	154	166	159	147	106	122
8	285	294	170	168	158	161	155	135	132	114
9	219	295	166	126	118	127	113	148	135	101

表 8: EXCITE の場合の数の分布 (00 から 99 まで)

計 15204 個	0	1	2	3	4	5	6	7	8	9
0	431	127	126	109	110	46	42	44	55	55
1	110	147	148	103	131	107	71	75	92	72
2	73	153	107	144	106	90	81	108	63	81
3	97	109	102	96	76	64	74	81	64	45
4	76	104	99	73	62	65	57	76	56	58
5	102	79	88	70	53	39	45	40	40	47
6	67	67	91	74	57	62	39	54	40	46
7	52	75	96	59	65	42	53	48	46	36
8	73	68	67	51	47	52	49	36	55	63
9	59	76	72	57	50	44	51	45	42	32





この上のグラフは、最手前、最左が00、最奥、最右が99となっている。表から0と1の出現回数が非常に多いことが分かる。また00や11など0や1を基にした数なども多いことが分かる。さらに3社ともグラフの形状は差異はあるものの、近いとみてよいだろう。このように人工的に作った数は、 $\sqrt{2}$ 、e、オイラー定数、 $\pi$ 、Log 2にくらべ偏りが大きい。逆にいえば、ある不明な文字や数字のデータについて、その中に現れる数字の頻度をカウントし、もしその分布が今調べたような分布に近ければ、そのデータは経済のデータである可能性が強いのであるといえよう。

## 0.8 反省

・プログラムは数の区切り方は前から一桁区切り、二桁区切り、三桁区切り、四桁区切りとしている。例えば二桁区切りについては、4142135...を41、42、13、5...と区切るやり方と4、14、21、35、...と区切るやり方の2種類があるが、私のやり方は前者のみを採用している。そのため数に偏りである可能性があることを加味せねばなるまい。ところが正規数の定義はその区切り方によらず定義されていることが驚くべきことである。

・ $\sqrt{2} = 1.4142135\dots$ の4,1,4,2,1,3...を独立で一様な確率変数列と仮定する。」としたが、明らかにこの仮定は偽である。

・研究当初から、このようなアプローチでは数の正規性を探ることはできないだろうということはわかっていた。しかし、実際に調べてみることはそれなりに価値があったと信じている。  
。・経済のデータだけでなく、様々な人工的に作られた文章などを調べてみたかった。しかし経済のデータ以外は非常に数字が少ないため、十分な量の数字を集めるには、さらに多量の元のデータが必要であったため実現できなかった。

参考書；

岩波数学辞典（正規数の定義についてのところ）  
カーニハン&リッチー C言語プログラム書

お世話になった方；(あいうえお順)

上山 大信先生  
柴田 達夫先生  
松本 眞先生  
若木 宏文先生

この卒業論文を書くにあたり、大変お世話になりました。