円周率πの小数表示に於ける桁数字の統計的考察

A note on the statistical analysis of sample digits of decimal digits of π

ネットワーク情報学部　田中　稔
School of Network and Information　Minoru TANAKA

Keywords: π, irrational number, e, transcendental number, correlation

1.　はじめに

3.14159…という数字は何を表すでしょうか？ほとんどの人は記憶にあるでしょう、これは小学校のころから親しみのあるπ，円周率の近似値である。あなたは円周率πの値を何桁まで覚えているだろうか。このπの値は無理数であり、十進小数で表記すると繰り返しのない無限小数であることが知られている。因みに、50桁までの円周率は

3.14159265358979323846264338327950288419716939937 51

となる。ところで、ここに最近購入した円周率πに関する本がある。

不思議な数のπ伝記（A.S.Posamentier&I.Lehmann, 松浦俊輔訳、日経BP社）（＊）

円周率πに関していろいろと詳しく書かれている。例えば、円周率にπの文字を本格的に使い始めたのは1700年代からでスイスの数学者レオンハルト・オイラーによる。3月14日はπの日といわれ、かの有名なアインシュタイン博士の誕生日でもある…。また、この本には紀元前2000年からのπの小数表記の計算年表が載っている。それによると、10年前（1995年）までは1千万桁程度であったが、現在では飛躍的に進歩して1兆2411億桁まで計算されているらしい（金田康正博士、2002年11月世界記録、約600時間)。さらに、金田博士は小数点以下の桁数字0から9までの数字の出現頻度（度数）分布を詳しく調べている。期待に反して、その分布は（今のところ）一様であるらしい。ここで次のような疑問がわいてくる。

問題：円周率πは無理数であるが、$\sqrt{2}$などと違い有理数係数の代数方程式の解（代数的数）ではない無理数（超越数transcendental number）であることが知られている（証明はリンデマン)。その他の超越数にはネピア数eがある（例えばSiegel [1]を参照)。そこで、ネピア数eも小数表記の数字の出現頻度に関してπと同じ性質をもつのだろうか。また、$\sqrt{2}$などの代数的無理数と超越数の間に小数表記の桁数字の出現性に関して何か違いがあるのだろうか。

本稿ではこの問題に関して特に統計的な性質に限定して考察する。そこで、今回私のパーソナルコンピュータ（CPU：Athlon 64　3200+、メモリー：DDR400 3GB）を使うことにした。果たしてパソコンでどの位まで計算できるのであろうか。計算ソフトに関しては、お馴染みの数式処理ソフトのMathematoica 5.1を使うことにした。結論から言うと、計算できたのは1000万桁までであった。さらに1100万桁にすると計算メモリーオーバーのサインが現れたのでやむなく中止したが、ソフト上の問題でなければメモリーを最大値4ギガバイトまで増設するなどの工夫で計算量を上げることが可能かもしれない。まず、得られたデータを使って前述の金田博士の結果を確認する。

(1) 小数表記数字の出現頻度（度数）分布は一様か（第2章）

更にここでは次のことを考える。小数点以下の数字を（桁を離散時間とする）時系列と見なす。小数第1位、2位、3位、…、t位の数字を順番に$X(1)$，$X(2)$，$X(3)$，…，$X(t)$と定義すると、これは離散値を取る離散時間の時系列と考えられる。従って通常の時系列解析の手法が使えるのではないか。そこで、円周率πとネピア数eの小数表記数字列における標本自己相関係数を計算し、数字の現れ方に関する何らかの規則性の存在の有無を検証してみる。

(2) 標本自己相関と無相関の検証（第3章）

また、πとeのそれぞれの数字列を2つの時系列と見なし、その数字の現れ方に相関（クロス相関）がないかを調べてみる。

(3) 円周率とネピア数の相互相関の検証（第4章）

第3章と第4章における時系列解析では、Mathematicaのアドインソフト Time Series を使って、特に標本自己相関、標本相互相関係数の計算、および独立性の検定などを行っている。

2.　小数表記数字の出現頻度分布は一様か

2.1　円周率πの少数点以下100万桁までの数字の出現頻度

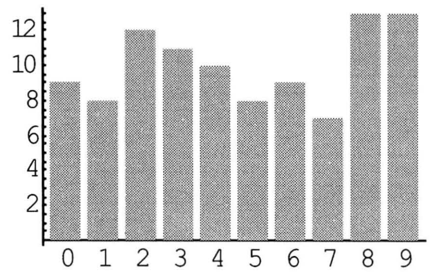
はじめに少数点以下1桁から100桁までの数字の列を求めてみよう。

```
nn=100;
pi=Drop[Characters[ToString[N[・,nn+1]]],2]
freq=Frequencies[pi]
BarChart[freq];
```

{1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 9, 7, 9, 3, 2, 3, 8, 4, 6, 2, 6, 4, 3, 3, 8, 3, 2, 7, 9, 5, 0, 2, 8, 8, 4, 1, 9, 7, 1, 6, 9, 3, 9, 9, 3, 7, 5, 1, 0, 5, 8, 2, 0, 9, 7, 4, 9, 4, 4, 5, 9, 2, 3, 0, 7, 8, 1, 6, 4, 0, 6, 2, 8, 6, 2, 0, 8, 9, 9, 8, 6, 2, 8, 0, 3, 4, 8, 2, 5, 3, 4, 2, 1, 1, 7, 0, 6, 8, 0 }

以下の2つの数字の組のリストは {度数, 数字} となっている。

{{9, 0}, {8, 1}, {12, 2}, {11, 3}, {10, 4}, {8, 5}, {9, 6}, {7, 7}, {13, 8}, {13, 9}}

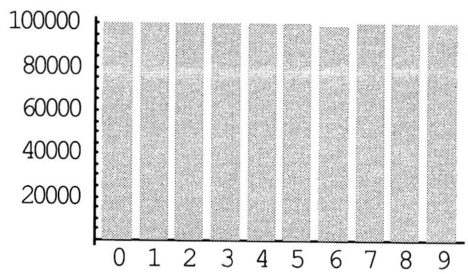


注意：この小数点以下100桁までの結果は、冒頭に引用した文献（*）にある数字の現れる回数（ページ133）と4カ所（数字0が8個，数字7が8個、数字8が12個、数字9が14個）で違った数値となっている。しかしながら、その他の千桁から1千万桁までの数値は同じであった。

同様にしてπの少数点以下百万桁までの頻度分布は以下のようになる。

```
nn=1000000;pi=Drop [Characters [ToString [N
[・, nn+1 ] ] ] , 2 ] ;
freq=Frequencies [pi]
BarChart [freq] ;
```

{{99959, 0}, {99757,1}, {100026, 2}, {100229, 3}, {100230, 4}, {100359 ,5}, {99548, 6}, {99800, 7}, {99985,8}, {100106, 9}}

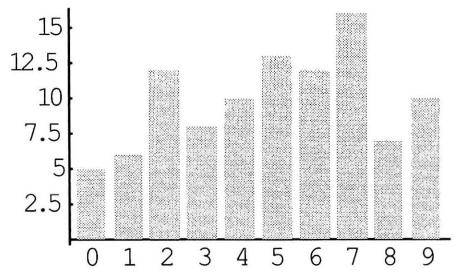


この段階で早くも金田博士の結果（1兆2411億桁までの数字の出現頻度の分布が一様である）が推察される。

2.2 ネピア数eの小数点以下100万桁までの数字の出現頻度

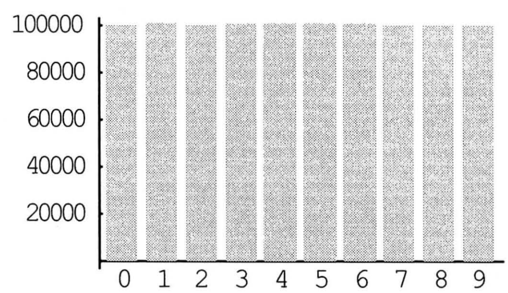
次にネピア数eの少数点以下100桁までの頻度分布を眺めてみよう。

{{5, 0}, {6, 1}, {12, 2}, {8, 3}, {10, 4}, {13, 5}, {12, 6}, {16, 7}, {7, 8}, {10, 9}}



この段階では円周率πの分布とはかなり異なっていることが判る。特に数字7が多く、0や1が少ない。そこで、さらに少数点以下百万桁までの頻度分布を求めてみる。

{{99425, 0}, {100132, 1}, {99845, 2}, {100228, 3}, {100389 ,4}, {100087, 5}, {100479, 6}, {99910, 7}, {99812, 8}, {99692, 9 }}



予想に反して、この結果も円周率πと同様に0から9までの数字の出現頻度には大きな差がみられず、やはり頻度分布は一様に感じられる。

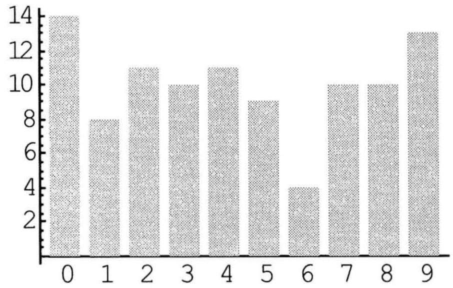
2.3 離散一様擬似乱数の100万個の数字の出現頻度

円周率やネピア数の0から9までの数字の出現頻度が一様分布に思われることから、実際の一様分布ではどのように数字が出現されるのかを調べてみる。Mathematica の組み込み関数Random を使って、擬似乱数のシミュレー

ションをしてみる。離散一様擬似乱数100個の頻度分布の一例：

```
nn=100;
rand100=Table[Random[Integer, {0, 9}], {nn}];
freq=Frequencies[rand100]
BarChart[freq] ;
```

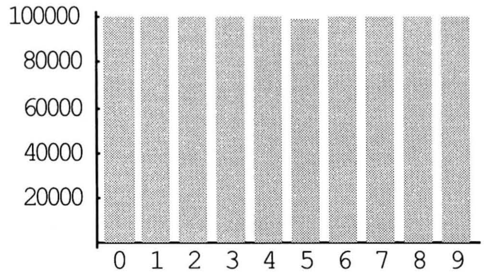
{{14, 0}, {8, 1}, {11, 2}, {10, 3}, {11, 4}, {9, 5}, {4, 6}, {10, 7}, {10, 8}, {13, 9}}



この結果は1つの例であって、シミュレーション実行のたびに異なるが、最小度数（4）と最大度数（14）の値などが参考になるであろう。さらに百万個の頻度分布の例としては以下のようになる。

```
nn=1000000;
rand100=Table [Random [Integer, {0,9}] , {nn}] ;
freq=Frequencies [rand100]
BarChart [freq] ;
```

{{100102, 0}, {100030, 1}, {100079, 2}, {100306, 3}, {99993, 4}, {99445, 5}, {100210, 6}, {99819, 7} ,{99912, 8}, {100104, 9}}



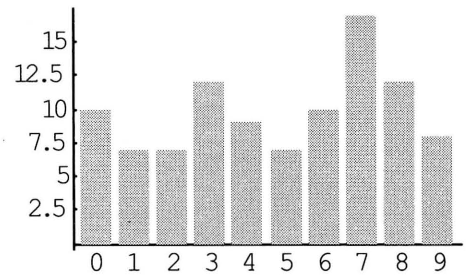
無理数の小数点表示の桁の数字と擬似乱数の数字列との違いに関して、出現頻度のグラフから判ることは数百、数千個のオーダーの数字列であればその頻度には偏りが見られるが、数百万のオーダー以上となると最早その違いはないことが明らかであろう。

2.4 代数的無理数の小数点以下100万桁までの数字の出現頻度

参考のため、円周率やネピア数以外の代数的無理数（$\sqrt{2}$, $\sqrt{3}$等）に関して調べてみる。以下は$\sqrt{2}$の場合である。

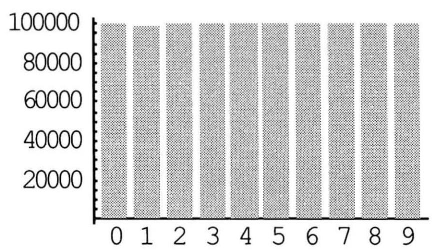
```
nn=100;
pi=Drop [Characters [ToString [N [√2, nn ] ] ] , 2 ] ;
freq=Frequencies [pi]
BarChart [freq] ;
```

{{10, 0},{ 7, 1}, {7, 2}, {12, 3}, {9, 4}, {7, 5}, {10, 6}, {17, 7}, {12, 8}, {8, 9 }}




```
nn=1000000;
pi=Drop [Characters [ToString [N [√2, nn ] ] ] , 2 ] ;
freq=Frequencies [pi]
BarChart [freq] ;
```

{{99814, 0}, {98924, 1}, {100436, 2}, {100190, 3}, {100024, 4}, {100155, 5}, {99886, 6}, {100008, 7}, {100441, 8}, {100121, 9}}



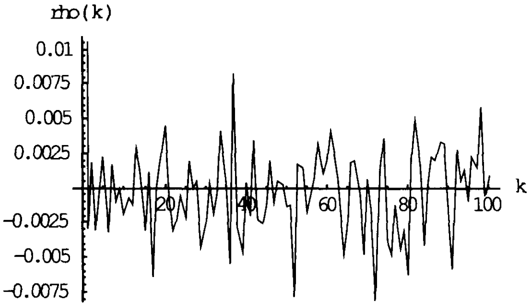
$\sqrt{2}$以外の無理数についてもこの結果と変わらなかったことを付け加えておく。無理数の小数表記における桁数字の出現頻度の分布から、ある無理数が超越数か代数的数かの判別をすることは不可能に思われる。

3. 標本自己相関と無相関の検証

3.1 円周率πの小数点数字の時系列の標本自己相関

無理数πの少数点数字列を離散値離散時間の時系列と見なしたとき、有意な自己相関が存在するか。また、2つの異なる無理数の少数点数字列の間に相互相関の関係が存在するか。

小数10万桁数字時系列の標本自己相関関数
(遅れk＝1〜100)のグラフ



特にラグ(遅れ) h = 1の場合について詳しく調べてみる。以下の表は遅れ1のペアー{X(t), X(t+1)}(総数は1千万－1個)の分布である。{0, 0}から{9, 9}までの100組の度数である。

遅れ1の組み合わせ{X(t), X(t+1)}の頻度分布

99662	100434	100065	100053	100221	100164	99768	99352	99902	99819
100058	99675	100196	99721	100248	100100	99502	100489	99740	99603
100169	99660	99931	99314	100239	99833	100617	100051	100289	100203
99641	100233	99809	100555	99743	99723	100058	100457	99992	99753
100217	100232	99991	100008	99915	100073	100394	99944	99994	100325
99972	99680	99952	99982	100168	100490	99952	100010	100304	99956
99432	99690	100080	100008	100298	99965	99905	100025	100292	99642
99584	99637	100207	99667	100311	100421	99686	100208	100202	100284
100816	100027	99664	100057	100028	99512	99823	99720	99781	100386
99889	100065	100411	100599	99922	100185	99632	99950	99318	100069

{最小値, 最大値}　{99314, 100816}
{平均値, 分散}　{100000., 91702. 4}

これよりどの組み合わせもほとんど同じ頻度で出現していることが分かる。従って、この状態が続くと仮定すれば、遅れ1の自己相関はゼロ(X(t)とX(t+1)は無相関、マルコフ性がない)と考えられる。

3.2 ネピア数eの小数点数字の時系列の標本自己相

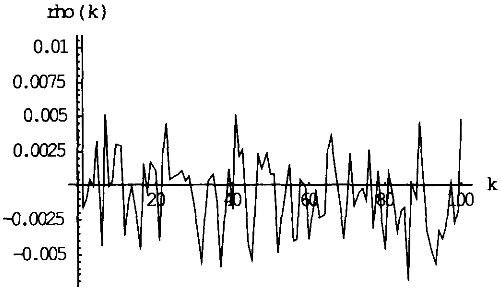
遅れ1の組み合わせ{X(t),X(t+1)}の頻度分布

100144	99853	99769	99670	100113	100139	99456	99833	99588	100089
99980	99887	99809	100647	99796	100155	100182	100369	99859	99900
99587	100231	99984	99970	100452	99964	99652	99625	99983	99719
99986	100185	99576	100669	99830	99832	100078	101063	99655	100829
99616	100017	99733	100097	100174	100228	99825	100327	100273	100029
99478	100137	100073	100234	99847	99912	100309	99996	99769	100145
100017	100468	100367	100240	99632	99306	99545	99822	100057	99420
100491	100115	99829	100246	99623	100464	99995	99718	100546	99796

```
{Min[t2],Max[t2]}
 {99306,101063}
 f2=Flatten[t2];Apply[Plus,f2]
 9999999
 {Mean[f2],VarianceMLE[f2]}//N
 {100000.,102585.}
```

3.3 離散一様擬似乱数の場合

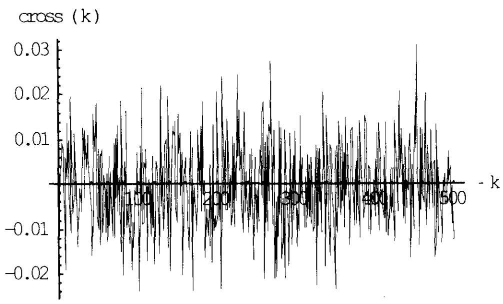
乱数10万個時系列の標本自己相関関数
(遅れk＝1〜100)のグラフ



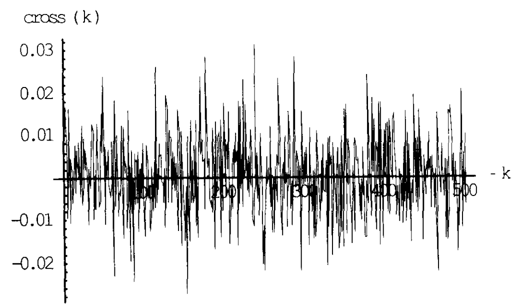
4. 円周率πとネピア数eの相互相関の検証

円周率とネピア数の小数表記の桁数 n ＝10000 の場合で、2つの数字列の標本相互相関係数を遅れk＝－100〜100までを計算する。

```
cross2=CorrelationFunction[toee200,topi200,nn/20];
plotcorr[cross2,AxesLabel?{"-k","cross(k)"}];
```




```
cross2=CorrelationFunction[topi200,toee200,nn/20];
plotcorr[cross2,AxesLabel?{"+k","cross(k)"}];
```



どの遅れの標本相互相関係数の値もその絶対値は、ほぼ $\frac{3}{\sqrt{n}}$ (上の例では絶対値が＝0.03)以内にあり、相関がない(母相互相関係数はゼロ)と推察される。従って、円周率とネピア数の小数表記における1万桁までの数字の列の間には相互相関はないと考えられる。小数表記の桁数を100

万桁まで増やしてもその結果は変わらなかった。(どうも離散値時系列の場合の「相関」の取り扱いに問題がありそうに思われる。)

5. まとめ

無理数は十進小数で表記すると「繰り返しのない」無限個の数字の列であるが、残念ながら我々はその全貌を垣間見ることは不可能である。これまでの結果をまとめてみると、ここで扱ったどの無理数に関しても、その1000万桁までの数字の列はほとんど(統計的に)ランダムな数列であり、その数字の出現性に関する規則(一様性以外の特徴)を見つけることは出来なかった。この状態でどこまでも無理数の小数表記における桁数字の列が無限に続くのであれば、数字の出現頻度に注目して「ある無理数が超越数か代数的数か」の判別をすることは不可能に思われる。「有限の世界」から「無限の世界」を覗くにはどうすればよいのだろうか。超越数の問題は院生の時代から関心のあったものだが、数理統計学の世界にも通じるところがあり奥深い。ただ今日ではコンピュータの発達やMathematicaなどの数理科学用ソフトの充実で、円周率πやネピア数eなどの超越数の問題が幾分身近になってきたことは確かであろう。ここでは時間の都合上、統計解析の簡単な手法でしか分析できなかったが、この問題を扱うにはもっと精密な手法が要求される。どうやら今回パンドラの箱を開けそうになったようだ。

参考文献

[1] C.L. Siegel, Transcendental Numbers, Princeton University Press, 1949.