

Fractal and Statistical Analysis on Digits of Irrational Numbers

Dejian Lai^{a,b}; Marius-F. Danca^c

April 17, 2006

^a Division of Biostatistics

University of Texas School of Public Health, RAS W1006

Houston, TX 77030, USA

Email: dejian.lai@uth.tmc.edu

^b Faculty of Statistics

Jiangxi University of Finance and Economics

Nanchang, Jiangxi, China

^c Department of Mathematics

Tehnofrig College

3400 Cluj-Napoca, Romania

Abstract

In this study, for quantifying and comparing the complexity of the digits of irrational numbers, we used the first one million digits to calculate the fractal dimensions of the digits of ten irrational numbers with long sequence of known digits via box-counting algorithm. The irrational numbers we studied are five algebraic numbers ($\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}$ and $\sqrt{7}$) and five transcendental numbers ($\pi, e, \log(2), \zeta(3)$ and Champernowne's constant). For statistical analysis, we performed 2000 repeated calculations for each number with segment of digits with length 200, 300, 400 and 500. The distributions of the estimated fractal dimensions seem normal by the histograms. Analysis of variance was used to test the equality of means for the fractal dimensions among the numbers and within the number for different digit lengths. The pattern of complexity of the digits in these numbers based on the estimated fractal dimensions is similar except that of Champernowne's constant.

1 Introduction

Many natural and social systems express complex, irregular patterns. One of the commonly used techniques in quantifying the complexity is based on estimating fractal dimensions.[1,2] In time series setting, fractal dimensions can be expressed in various forms such as Korcak exponent and Hurst coefficient. Multi-dimensional scaling such as Korcak exponent was estimated on Japanese earthquake time series and Chinese earthquake time series.[2,3] The Hurst effect was applied in monitoring clinical trials.[4,5] Fractal analysis has been applied in many diversified fields.[2]

In this article, we report our estimation and statistical analysis of the fractal dimension via box counting algorithm on the complexity of the digits of several widely used algebraic and transcendental irrational numbers.[6,7] As we know that the real numbers can be classified into rational numbers and irrational numbers. Any rational number r can be expressed as a ratio of two integers p and q as $r = p/q$. The rest of the real numbers besides rational numbers are called irrational numbers. From a Lebesgue measure point of view, there are more irrational numbers than rational numbers. For example, in the interval of $[0,1]$, the Lebesgue measure of rational numbers is 0, whereas it is 1 for the irrational numbers. It is well known that the digits

of rational numbers will keep repeating their patterns as a finite number of digits. However, the digits of irrational numbers show very complicated structures. Further, irrational numbers can be divided into algebraic and transcendental. A transcendental number is any irrational number that is not a root of any polynomial with integer coefficients. An irrational number which is a root of a polynomial with integer coefficients is called algebraic number. Almost all irrational numbers are transcendental, while all transcendental numbers are irrational.

The number of decimal digits for any irrational number is infinite. It is a great challenge to mathematicians and computer scientists to find out a long sequence of accurate digits for irrational numbers. In our analysis of the complexity of the digit patterns of irrational numbers, we create “time series” from the digits by treating the position as the “time” index and the digit at the position t as the observed value of the “time series” $Z(t)$, where $Z(t)$ representing the digit at position t for a given irrational number. We studied five algebraic numbers ($\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}$ and $\sqrt{7}$) and five transcendental numbers (π , e , $\zeta(3)$, $\log(2)$ and Champernowne’s constant). The dynamics of the digits of some irrational numbers have been extensively studied. For example, digits of the well-known irrational number π (the ratio of the circumference of a circle to its diameter) were suggested as natural random numbers.[8] It is generally not easy to show one number is irrational and event harder to establish its transcendence. Since the rational numbers and the algebraic numbers are countable, hence, the set of transcendent number is uncountable. Hence there are much more transcendental numbers than algebraic irrational numbers. However it is very difficult to prove that a particular number is transcendental and only few transcendental numbers are known After the irrationality of $\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}$ and $\sqrt{7}$ is shown, these numbers are obviously algebraic. The irrationality of π is proved by Lambert in 1771 and Lindermann established its transcendence in 1882 (Dodge 1996). Besides π , e is another widely used transcendental irrational number that can be written as $e = \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$. $\zeta(3)$ can be expressed as $\zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3}$. In our study, $\log(2)$ is the value with base e . Champernowne’s constant C_{10} was constructed in 1933

by concatenating the positive integers to the right of a decimal point.[9] Analytically, $C_b = \sum_{n=1}^{\infty} \frac{C_{b,n}}{b^{S_{n,b}}}$, where $C_{b,n} = \sum_{k=b^{n-1}}^{b^n-1} kb^{-n[k-(b^{n-1}-1)]}$ and $\sum_{k=0}^{n-1} k(b-1)b^{k-1}$, $[\cdot]$ is a floor function such that $[3.6] = 3$. For example, $C_{10}=0.12345678910111213141516171819202122232425\dots$ We downed one million digits from the websites for our fractal and statistical analysis on the digits of the ten irrational numbers.[10]

2 Box Counting Algorithm

There are many ways to define the fractal dimension of a subset in Euclidean space. In this article, we treat the digits of the irrational numbers as a “time series” along the positions in the two dimensional Euclidean vector space and estimate the fractal dimension through box counting algorithm.

Definition 1 *The box dimension of a subset X is defined as*

$$D = \lim_{\Delta s \rightarrow 0} \frac{\log N(\Delta s)}{-\log(\Delta s)}, \tag{1}$$

where $N(\Delta s)$ denotes the number of boxes in a grid of linear scale Δs which meets X .

A convenient graphical representation of the digits of an irrational number is to plot the digits versus digit’s position, which is a discrete transformation of α . We call the series of digits in this paper as “time series”. The graph of this time series could be considered as a subset in the two dimensional Euclidean space R^2 . The first 500 digits of $\sqrt{2}$ are plotted in Figure 1. The one million digits of the irrational numbers were downloaded from websites.[10]

Figure 1 about Here

In our calculation, we meshed grid systems with four different values (2, 3, 4, 5)/length of digits for the Δs . The box dimension was then estimated via regressing $\log N(\Delta s)$ on $-\log(\Delta s)$. The estimation of the fractal dimensions was performed for each of the ten numbers

at length of 200, 300, 400 and 500 digits 2000 times. A typical regression linear is shown in Figure 2 for the first 500 digits of $\sqrt{2}$.

Figure 2 about Here

The program of computing the fractal dimension was written in Turbo Pascal. The actual computation was performed on a 500 MHZ PC. A slightly different version of box-counting algorithm was used by Chatterjee and Yilmaz in selecting the box length for random series.[7] Theoretical justifications of box counting algorithm is reviewed by Barnsley.[11]

3 Statistical Analysis

Using the box counting method, we computed the fractal dimensions of digits of five algebraic numbers ($\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{6}$ and $\sqrt{7}$) and five transcendental numbers ($\pi, e, \zeta(3), \log(2)$ and Champernowne's constant). Our calculations were performed for digits of length 200, 300, 400 and 500. For a proper statistical analysis, we repeated our computation 2000 times consecutively along the digit positions. For example, if the length of the segment is 500, our computation used $500 \times 2000 = 1000000$ digits for each of the irrational numbers. Similar computations were performed for other numbers with 200, 300 and 400 digits in each segment.

We have compared the histograms of the five algebraic numbers and the five transcendental numbers. These histograms show that the estimated fractal dimensions are quite normally distributed except the estimate of the fractal dimensions for the digits of the Champernowne's constant, which are slightly differ from histograms of normal random variables. The histograms and other plots were generated using S-plus.[12] In Figures 3 and 4, we show the histograms of $\sqrt{2}$ and the Champernowne's constant. Other histograms are available upon request.

Figures 3 and 4 about Here

Results in Table 1 show a slight increase of the means of the estimated fractal dimensions as the length of the segment increases all the irrational numbers considered except Champernowne's constant. The variances of the estimates decrease as the length of the digits increases, which indicates that the estimates converge to their limits of Hausdorff dimensions of each of the digit series.

Table 1 about Here

In comparing the means of the estimates for various segment lengths within each of the ten irrational numbers, we conducted analysis of variance (ANOVA) on the means using SAS.[13] Although there were significant differences among the means, only results from the segment length of 500 were statistically significant different with a p -value less than 0.05 from that of other lengths for the nine irrational numbers except Champernowne's constant. It is clear that the digit structure of Champernowne's constant is significant different from all other nine irrational numbers using the box counting fractal dimension. The p -values of the pairwise comparisons within each number are given in Table 2. The p -value is the upper bound of probabilities that one would reject the null hypothesis (equality of means) when it is true. A commonly used threshold value for the p -value is 0.05.[14]

Table 2 about Here

An analysis of variance was also applied in comparing the means of the estimated fractal dimensions among the irrational numbers for the four different length of 200, 300, 400 and 500. For digits with length 200, there was no significant difference for the means of the estimated fractal dimensions among the ten irrational numbers. For the digits with length greater than 300, we observed the estimated fractal dimensions of Champernowne's constant were significantly different from the fractal dimensions estimated from other nine irrational numbers ($p < 0.0001$). The p values from these pairwise comparisons and the individual

estimates of the fractal dimensions for each length and irrational number are available upon request.

4 Concluding Remarks

Based on our study, we demonstrated that the fractal dimensions calculated by box counting algorithm for the digits of the nine of the ten irrational numbers were statistically equivalent although the precision resulting from short digit series may be low. The structure of the digits of the Champernowne's constant showed significant difference comparing to other nine irrational numbers. It is worth noting that, for these ten irrational numbers, only Champernowne's constant is fabricated with intentional arrangement of the digits whereas the digits from other name irrational numbers were placed intrinsically.

There seems no clear distinguish for fractal dimensions of the digits between the algebraic numbers and transcendental numbers except Champernowne's constant. The fractal dimensions of the digits of Champernowne's constant estimated from reasonable length (> 300) were significantly difference from the five algebraic numbers and the other four transcendental numbers.

For each number, our analysis of variance does not take into account the possible correlations of the estimated fractal dimensions. Although there are more transcendental numbers than algebraic numbers, there are relatively few numbers were shown to be transcendental. Further, we can only find the above five transcendental numbers with enough digits to allow us to compute fractal dimensions for statistical analysis.

The apparent normal distribution of the estimates of the fractal dimensions for the nine of "natural" irrational numbers (except Champernowne's constant) is worthing further theoretical studies. Notice that the box counting fractal dimension was a estimate of the slope from the regression line of the number of boxes intersecting with the connecting line between the digits. These numbers increase as the box length width decreases. It is expected that

the number of the boxes would skew to the left with smaller Δs if there digits appeared randomly in the two dimensional Euclidean space. Hence, the number of boxes $N(\Delta s)$ may be approximated by a lognormal distribution. That is, $\log N(\Delta s)$ is approximately normally distributed. A rigorous proof of these approximations is beyond the scope of this paper.

The statistical properties of the digits of π were well studied.[8] The digits from irrational numbers possess complicate patterns. These patterns may be useful in many different fields such as in information sciences and in measuring health.[15,16] Statistical analysis of the fractal dimensions is a way of quantifying and comparing these complex natural phenomena.

References

1. Mandelbrot, B. B. *Fractals: Form, Chance and Dimension*. (Freeman, San Francisco, 1977).
2. Hastings, H. M. & Sugihara, G. *Fractals: A User's Guide for the Natural Sciences*. (Oxford University Press, New York, 1995).
3. Lai, D. Statistical and fractal analysis of Chinese earthquake time series. *Fractals*. **8**, 123-127 (2000).
4. Lai, D., Davis, B. R., Hardy, R. J. Fractional Brownian motion and clinical trials. *Journal of Applied Statistics*. **27**, 103-108 (2000).
5. Lai, D.: Estimating the Hurst effect and its applications in monitoring clinical trials. *Computational Statistics and Data Analysis*. **45**, 549-562 (2004).
6. Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications*. (Springer, New York, 1988).
7. Chatterjee, S. & Yilmaz, M. Use of estimated fractal dimension in model identification for time series. *Journal of Statistical Computation and Simulation*. **41**, 129-141 (1992).
8. Dodge, Y. A natural random number generator. *International Statistical Review*. **64**, 329-344 (1996).
9. Champernowne, D. G. The construction of decimals normal in the scale of ten. *Journal of London Mathematical Society*. **8**, 254-260 (1933).
10. <http://antwrp.gsfc.nasa.gov/htmltest/> and <http://ja0hxxv.calico.jp/pai/estart.html>. Accessed in January 2005.

11. Barnsley, M. F. *Fractals Everywhere*. (Academic Press, Boston, 1988).
12. Insightful. *S-plus 6.2 Professional Edition*. (Insightful, Seattle, 2003)
13. SAS Institute. *SAS System for Windows 8.2* . (SAS Institute, Cary, 2003).
14. Lehmann, E. L. & Romano, J. P. 2004. *Testing Statistical Hypotheses*, 3rd ed. (Springer, New York, 2004).
15. Pincus, S. & Singer, B. H. A recipe for randomness. *Proceeding of the National Academy of Science of the United States of America*. **91**, 10367-10372 (1998).
16. Portsen, N., Hollingdal, M., Juhl, C., Butler, P., Veldhuis, J. D. & Schmitz, O. Pulsatile insulin secretion: Detection, regulation, and role in diabetes. *Diabetes*, **51**, S245-S254 (2002).

Acknowledgments

Thanks are due to Emil Burton for discussions that lead to the work and writing of this article.

Table 1: Summary Statistics of the Estimated Fractal Dimensions of the Digits with Length 200, 300, 400 and 500 of Irrational Numbers Using 2000 Segments.

Number	200		300		400		500	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\sqrt{2}$	1.3718	0.0293	1.3752	0.0246	1.3761	0.0208	1.3766	0.0184
$\sqrt{3}$	1.3729	0.0291	1.3756	0.0247	1.3765	0.0206	1.3768	0.0184
$\sqrt{5}$	1.3725	0.0291	1.3758	0.0244	1.3765	0.0205	1.3766	0.0182
$\sqrt{6}$	1.3739	0.0287	1.3757	0.0238	1.3771	0.0195	1.3772	0.0180
$\sqrt{7}$	1.3730	0.0295	1.3758	0.0234	1.3766	0.0203	1.3768	0.0182
π	1.3742	0.0278	1.3762	0.0233	1.3763	0.0206	1.3767	0.0183
e	1.3727	0.0281	1.3738	0.0231	1.3772	0.0204	1.3777	0.0185
$\zeta(3)$	1.3728	0.0292	1.3756	0.0233	1.3771	0.0201	1.3769	0.0178
$\log(2)$	1.3731	0.0292	1.3763	0.0227	1.3773	0.0200	1.3775	0.0179
Champernowne's Constant	1.3714	0.1023	1.3625	0.0684	1.3479	0.0884	1.3573	0.0868

Table 2: The p-values of Pairwise Comparisons of the Means of Various Segment Lengths:
(a) 200 vs 300; (b) 200 vs 400; (c) 200 vs 500; (d) 300 vs 400; (e) 300 vs 500; (f) 400 vs 500.

	(a)	(b)	(c)	(d)	(e)	(f)
$\sqrt{2}$	0.4808	0.0624	<0.0001	0.2467	<0.0001	<0.0001
$\sqrt{3}$	0.6574	0.0936	<0.0001	0.2174	<0.0001	<0.0001
$\sqrt{5}$	0.8928	0.2868	<0.0001	0.3521	<0.0001	<0.0001
$\sqrt{6}$	0.8335	0.0301	<0.0001	0.0502	<0.0001	<0.0001
$\sqrt{7}$	0.7975	0.1969	<0.0001	0.3012	<0.0001	<0.0001
π	0.5862	0.4290	0.0004	0.8052	0.0029	0.0029
e	0.4964	0.0061	<0.0001	0.0393	<0.0001	<0.0001
$\zeta(3)$	0.7398	0.0807	<0.0001	0.0367	<0.0001	<0.0001
$\log(2)$	0.7862	0.0925	<0.0001	0.1581	<0.0001	<0.0001
Champernowne's Constant	0.0006	0.0587	<0.0001	<0.0001	<0.0001	<0.0001

Figure Captions

Figure 1: The Plot of the First 500 Digits of $\sqrt{2}$.

Figure 2: The Regression Line Based on the Box Counting Algorithm for Estimating the Fractal Dimension for the First 500 Digits of $\sqrt{2}$.

Figure 3: The Histograms of the Estimates of the Fractal Dimensions for the Digits of $\sqrt{2}$ with Length: (a) 200, (b) 300, (c) 400, (d) 500.

Figure 4: The Histograms of the Estimates of the Fractal Dimensions for the Digits of Champernowne's Constant with Length: (a) 200, (b) 300, (c) 400, (d) 500.

Figure 1:

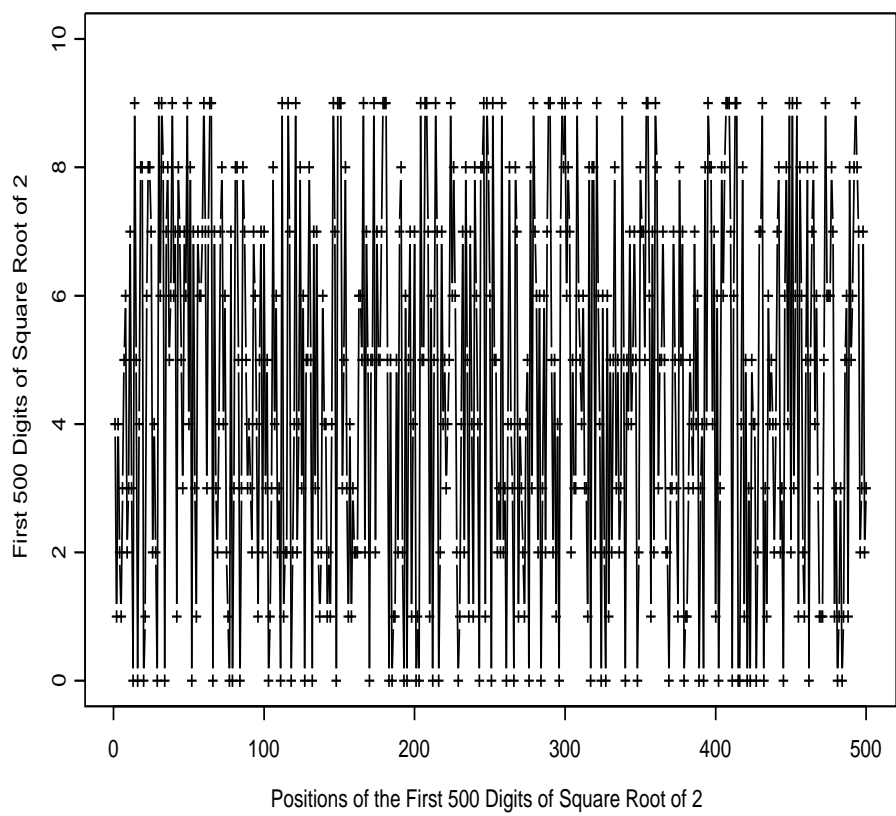


Figure 2:

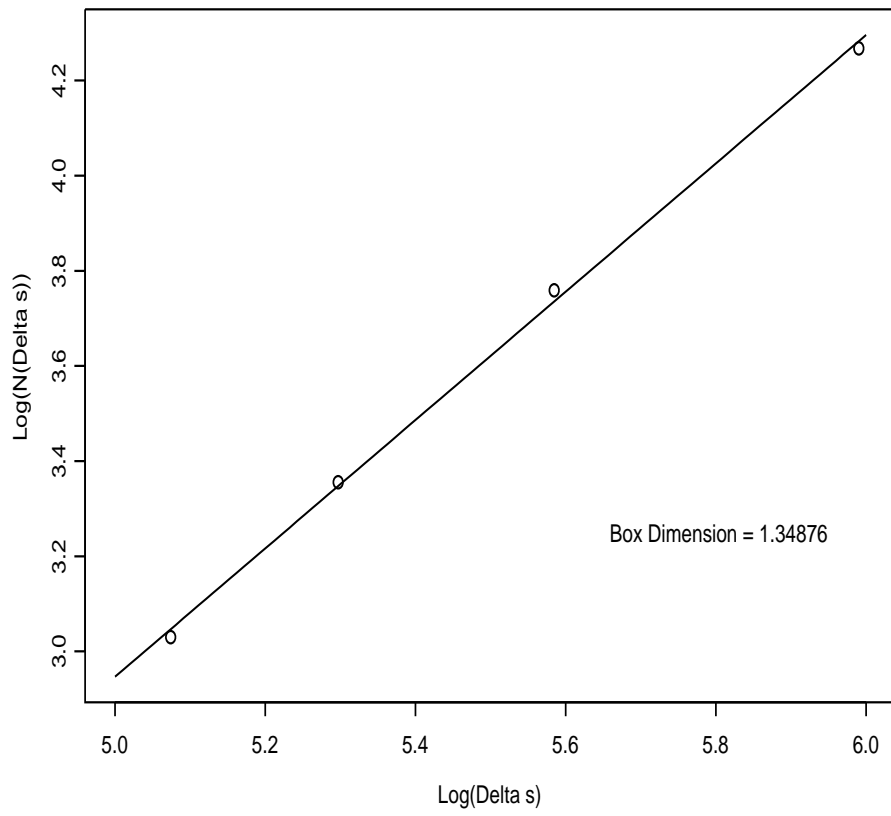


Figure 3:

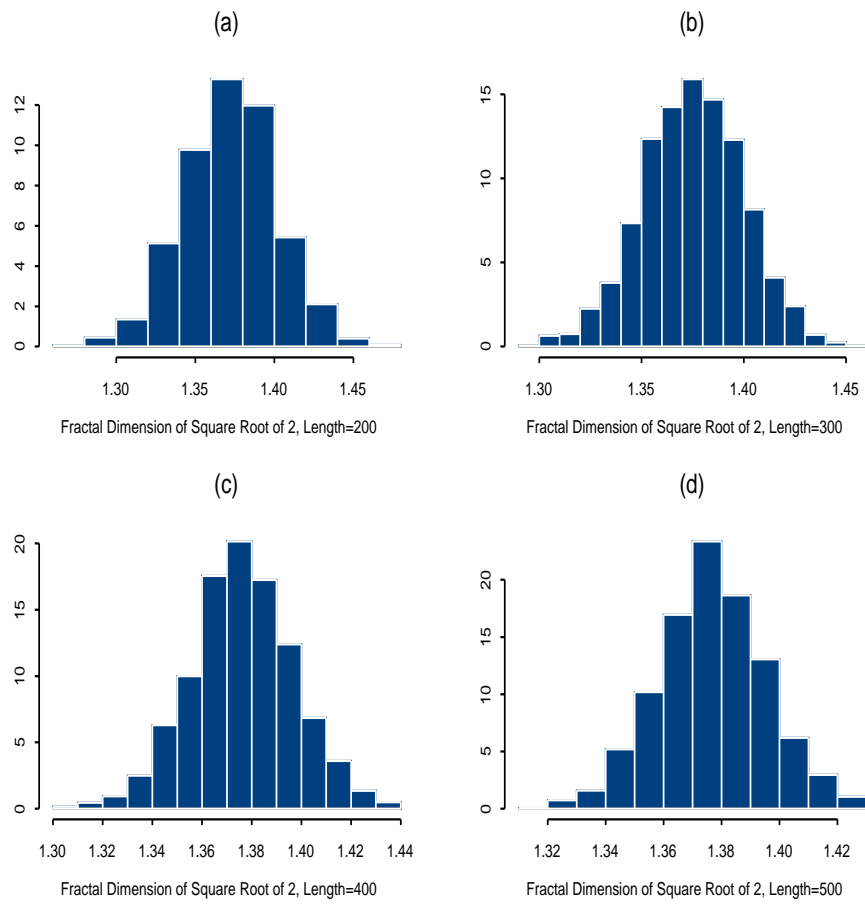


Figure 4:

