



The Possible Structure in the Distribution of Decimals in the Euler's Number Expansion

Alexander Bolotin*¹

¹Department of Computer Science, Ben-Gurion University of the Negev, Beersheba, Israel.

Article Information

DOI: 10.9734/BJMCS/2014/13537

Editor(s):

(1) Sergio Serrano, Department of Applied Mathematics, University of Zaragoza, Spain.

Reviewers:

(1) Anonymous, Kunming University of Science and Technology, China.
(2) Antonio Felix Flores Rodrigues, University of the Azores, Campus of Angra do Heroismo, Portugal.

Peer review History: <http://www.sciencedomain.org/review-history.php?iid=669id=6aid=6297>

**Original Research
Article**

Received: 21 August 2014
Accepted: 19 September 2014
Published: 01 October 2014

Abstract

This short paper continues to discuss randomness of the Euler's number e digits in decimal expansion. To analyze such randomness statistically, we exploit fixed-effects and random-effects Poisson panel-data models. The results of the regression models reveal the presence of some structure in the distribution of e decimals.

Keywords: Euler's number, Decimal expansion, Randomness, Poisson panel-data models.

1 Introduction

As it is well known, the absolutely exact value of the Euler's number e cannot be computed (given that – according to the Lindemann-Weierstrass theorem [1] – the number e is not algebraic); therefore, the representation of this number in any base would never end and would never settle into a permanent repeating pattern. It is conjectured that e is normal, meaning that when e is expressed in, say, decimal base

$$e = 2.718281828459045235360287 \dots = 2.d_1d_2 \dots d_i \dots d_{N-1}d_N \dots ,$$

*Corresponding author: E-mail: arkady_bolotin@hotmail.com

the possible decimal digits in the Euler's number d_i occur with equal probability

$$P(d_i = n) = P(d_j = n) \quad n \in [0, 9], i \neq j$$

in any sequence of given length. Albeit the decimal digits d_i indeed seem to be distributed randomly, no proof of such randomness has been found yet.

With no mathematically rigorous proof in sight, the main method of attacks on e randomness becomes statistical analysis. Many statistical tests have been used to determine to what extent the digits d_i are randomly arranged (see for example papers [2, 3], to name a few). Despite an apparent diversity of the exploited tests, they are all based on the same idea originated from works [4] and [5] of Marsaglia: These tests are to examine if the digits d_i , whether viewed one at a time, or in segments or, more generally, as arguments in complicated functions, can be reasonably considered to have arisen as the output of those same functions applied to a sequence of independent identically distributed (i.i.d.) random digits.

Even as such tests proved to be quite powerful (that is, they would be difficult to pass had the distribution of the digits d_i been not random), they, nevertheless, could be eluded if there were possible "structure" in a sequence of d_i . Explicitly, if, for example, a particular sequence $S_h = \{d_h, \dots, d_{h+km}, \dots, d_{h+lm}\}$ ($0 \leq k \leq l$) has the probability $P(d_{h+km} = n)$, which is greater (or less) than the corresponding probability $P(d_{j+km} = n)$ for the analogous disjoint $h \neq j$ sequence $S_j = \{d_j, \dots, d_{j+km}, \dots, d_{j+lm}\}$, then those random tests might not detect this difference when testing the ordered sequence $S = \{d_1, \dots, d_i, \dots, d_N\}$ that includes the elements of both sequences S_h and S_j .

On the other hand, panel data analysis (also known as cross-sectional time-series analysis) looking at the sequence $S_h = \{d_h, \dots, d_{h+km}, \dots, d_{h+lm}\}$ (that is, at the h^{th} panel unit) on more than one occasion $h = 1, 2, \dots$ would possibly be capable of detecting such change in the probability (for a full and systematic discussion of panel-data models, see papers [6, 7, 8]). Surprisingly panel data analysis was never done for the analysis of e randomness.

Thus, the aim of this short paper is to analyze the distribution of the digits d_i in the Euler's number e using fixed-effects and random-effects Poisson panel-data models.

The paper is organized as follows. First, a brief explanation of Poisson panel-data models will be presented; then the description of the panel data analysis will follow. The results of the analysis and the interpretation will conclude the paper.

2 Poisson Panel-data Models

Poisson panel-data regression fits via maximum likelihood the model

$$P(D_{ht} = d_{ht} \mid \mathbf{x}_{ht}) = F(d_{ht}, \mathbf{x}_{ht}\beta + \nu_h)$$

for $h = 1, 2, \dots$ panels, where $t = 1, 2, \dots, T_h$ is the object's current number in the h^{th} panel, \mathbf{x}_{ht} are covariates, β are their regression coefficients. If X_{ht} is the exposure, the expected number of inci-

dents, d_{ht} , will be

$$d_{ht} = \exp(\ln X_{ht} + \mathbf{x}_{ht}\beta + \nu_h) \quad .$$

In the standard random-effects model, ν_h is assumed to be i.i.d. such that $\exp(\nu_h)$ is gamma with mean one and variance α , which is estimated from the data. For a more detailed review of Poisson panel-data regression models see [9, 10, 11].

3 Description of the Statistical Analysis

In order to construct our statistical dataset, we use as the dependent variable $D = \{d_1, \dots, d_i, \dots, d_{3096}\}$ the first 3,096 decimal digits of e computed in the paper [12] by the famous series expansion given by Euler. As the independent variable of the dataset $X = \{1, \dots, x_i, \dots, 3096\}$, we use the place of a given decimal d_i (counted from the decimal separator) in the expansion of e in base ten. Using the variable X we generate four additional variables X_3, X_2, X_1 and X_0 in the following way:

$$\begin{aligned} X_3 &= \left\lfloor \frac{X}{10^3} \right\rfloor \quad , \\ X_2 &= \left\lfloor \frac{X}{10^2} \right\rfloor - X_3 \cdot 10 \quad , \\ X_1 &= \left\lfloor \frac{X}{10} \right\rfloor - X_3 \cdot 10^2 - X_2 \cdot 10 \quad , \\ X_0 &= X - X_3 \cdot 10^3 - X_2 \cdot 10^2 - X_1 \cdot 10 \quad , \end{aligned}$$

where $\lfloor \cdot \rfloor$ denotes the floor function of the variable X ; thus, if X is equal to, say, 1324, the generated variables X_3, X_2, X_1 and X_0 will be equal 1, 3, 2 and 4, respectively. It is clear that within the constructed dataset the span of the generated variables is $X_2, X_1, X_0 \in [0, 9], X_3 \in [0, 3]$.

To declare the data to be a panel, we identify the variable X_0 as the panel unit identification. Accordingly, the zero panel unit is the sequence $S_0 = \{d_{10}, d_{20}, \dots, d_{3090}\}$ of the decimal digits in e whose sequential number (i.e., its place) ends with '0', the 1st panel unit is the sequence $S_1 = \{d_1, d_{11}, \dots, d_{3091}\}$ of the decimals d_i whose sequential number ends with '1', and so on.

In this way, we get data on e decimals d_i (which we will consider as the numbers of some hypothetical "incidents") for ten different panel units $h = 0, 1, \dots, 9$. Our measure of "exposure" is X – the position of a digit from the decimal separator, and in our models we assume that the exponentiated random effects are distributed as gamma with mean one and variance α .

We wish to analyze whether the "incident" rate is affected by the variables X_3, X_2 and X_1 . Clearly, if the answer were 'yes', it would indicate the presence of some structure in the distribution of e decimals d_i .

4 The Results of the Analysis and the Interpretation

In the Tables 1 and 2, we present the output of two different Poisson panel-data regression models (implemented using the commercially available statistical software Stata, version 12.1, on a PC equipped with the Intel Core i7 3.1 GHz processor). The Table 1 also includes a likelihood-ratio test of $\alpha = 0$, which compares the panel estimator with the pooled (Poisson) estimator.

Table 1: **Random-effects Poisson regression of e decimals**

| <i>Variables in model</i> | <i>Coefficient</i> | <i>p-value</i> |
|---------------------------|--------------------|----------------|
| X_3 | -0.871779 | 0.000 |
| X_2 | -0.114161 | 0.000 |
| X_1 | -0.114161 | 0.000 |
| $\ln X$ | 1 | (exposure) |
| α | 0.000861 | |

Likelihood-ratio test of $\alpha = 0$: $\chi^2 = 4.07$ at $p = 0.022$

Table 2: **Conditional fixed-effects Poisson regression of e decimals**

| <i>Variables in model</i> | <i>Coefficient</i> | <i>p-value</i> |
|---------------------------|--------------------|----------------|
| X_3 | -0.8718306 | 0.000 |
| X_2 | -0.1141646 | 0.000 |
| X_1 | -0.010065 | 0.001 |
| $\ln X$ | 1 | (exposure) |

Table 3: **Comparing two sequences S_A and S_B of e decimals**

| <i>Statistics for sequences of decimals</i> | <i>S_A: decimals whose sequential numbers start with '00'</i> | <i>S_B: decimals whose sequential numbers start with '29'</i> |
|---|--|--|
| Number of decimals | 99 | 100 |
| Median | 5 | 4 |
| Mean | 4.939394 | 4.4 |
| Standard Deviation | 2.614005 | 3.117238 |
| Variance | 6.833024 | 9.717172 |
| Skewness | -0.1815752 | 0.0486636 |
| Kurtosis | 2.003474 | 1.651921 |

As it can be readily seen, both of these models with the almost identical efficiency show that the “incident” rate for the decimals d_i is significantly different for all the covariates, but especially for X_3 and X_2 .

To illustrate the results of the regression models, in the Table 3 we compare two sequence S_A and S_B of the decimals d_i whose sequential numbers $X = X_3X_2X_1X_0$ start with '00' and '29', respectively.

As it is supposed to be in accordance with the results of the regression models, the decimals of the sequence S_A turn out to be higher on average than those of the sequence S_B .

5 Conclusion

So, in conclusion we can say that in all probability (or at least for the first few thousand of digits) the distribution of decimals d_i in the Euler's number e is structured and thus not random.

Unfortunately, the statistical evidence against randomness of the decimal expansion of the Euler's number e , which we have demonstrated in this paper, leaves open the question whether the observed structured sequences of decimals d_i are related to the Euler's number itself or whether they are a product of arithmetical operations of approximation of that number (along the lines of the idea which was put forward in the work of Rodrigues and Martins, see [13]). But, in fact, this is the limitation of any statistical approach to e randomness since such an approach ignores the a priori knowledge that the decimals d_i of the Euler's number are the result of a computation and asks only whether the observed series of d_i behaves, within reasonable statistical norms, as though it were the realization of a sequence of i.i.d. random variables with discrete uniform distribution on $\{0, 1, \dots, 9\}$.

Acknowledgments

The author wishes to thank Dr. Arkady Bolotin of the Public Health Department, Ben-Gurion University of the Negev, for his statistical support and stimulating comments that made possible to compose this paper and two anonymous reviewers for their motivating remarks and invaluable help in improving the paper.

Competing Interests

The author declares that no competing interests exist.

References

- [1] Baker A. Transcendental Number Theory. Cambridge University Press; 1975.
- [2] Marsaglia G. On the Randomness of Pi and Other Decimal Expansions. *Inter. Stat.* 2005;5.
- [3] Shanks D, Wrench JW. Calculation of e to 100,000 Decimals. *Math. Comput.* 1969;23:679-680.
- [4] Marsaglia G. The Marsaglia Random Number CDROM, with The Diehard Battery of Tests of Randomness. 1985. Accessed September 5, 2014. Available: <http://www.cs.hku.hk/diehard/>
- [5] Marsaglia G, Tsang WW. Some difficult-to-pass tests of randomness. *Journal of Statistical Software.* 2002;7(3).

- [6] Baltagi BH. Econometric Analysis of Panel Data. 4th ed. New York: Wiley; 2008.
- [7] Hsiao C. Analysis of Panel Data. 2nd ed. New York: Cambridge University Press; 2003.
- [8] Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, MA: MIT Press; 2010.
- [9] Allison PD. Fixed Effects Regression Models. Newbury Park, CA: Sage; 2009.
- [10] Hardin JW, Hilbe JM. Generalized Linear Models and Extensions. 2nd ed. College Station, TX: Stata Press; 2007.
- [11] Cameron AC, Trivedi PK. Regression Analysis of Count Data. Cambridge: Cambridge University Press; 1998.
- [12] Sale A. The Calculation of e to many Significant Digits. Computing Journal. 1968;1:229-230.
- [13] Rodrigues A, Martins N. Numerical Uncertainty and Its Implications. Journal of Applied Mathematics and Physics. 2014;2:33-44.

©2014 Alexander Bolotin; This is an Open Access article distributed under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/3.0>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

www.sciencedomain.org/review-history.php?iid=669&id=6&aid=6297