

# Comment on the paper “A New Method for Symbolic Sequences Analysis. An Application to Long Sequences”

Marek Wolf

*Cardinal Stefan Wyszyński University  
Faculty of Mathematics and Natural Sciences. College of Sciences  
ul. Wóycickiego 1/3, Auditorium Maximum, (room 113)  
PL-01-938 Warsaw, Poland  
e-mail: m.wolf@uksw.edu.pl*

Received: 25 March 2015; accepted: 08 May 2015; published online: 06 June 2015

**Abstract:** We discuss several drawbacks in the recent paper [1] concerning some statements about the Champernowne number.

In the recent issue of CMST there appeared an interesting paper [1] by B. Kozarzewski on the new method for symbolic sequences analysis. This method was tested on several long sequences, in particular on the digits of the so called “Champernowne number”, which we will denote below as  $C_{10}$ . Unfortunately, some of the statements about this constant were not correct. The constant  $C_{10}$  was proposed by 21-year old David Gawen Champernowne in 1933 in the paper [2] as a concatenation after the decimal point of consecutive natural numbers expressed in the base of ten:  $C_{10} = 0.12345678910111213\dots$ . E. Borel had proved in 1909 [3] that almost all real numbers are normal and the first explicit example of normal number was  $C_{10}$ . On page 98 of [1], right column, it is written about  $C_{10}$  “The number is assumed to be transcendental”. In fact, the number is transcendental as it follows from the more general theorem proved by K. Mahler in [4]. Mahler has proved the following theorem: Let  $f(k)$  be a nonconstant polynomial with integer values for integer  $k$ , which is positive for  $k > 1$  and tends to  $\infty$  with  $k \rightarrow \infty$ . Then let  $\sigma$  denote the decimal fraction that arises if behind the decimal comma in sequence one after one the natural numbers  $f(1), f(2), f(3), \dots$  are written down, e.g.  $f$  be the polynomial  $f(k) = \frac{1}{2}(k^2 + k)$  leads to the fraction  $\sigma$  of the form in decimal base:

0.1 3 6 10 15 21 28 36 45 55 66 78 91 105 120

Then the theorem asserts that all  $\sigma$  constructed in this way are transcendental. Taking as a polynomial the linear function  $f(x) = x$  gives the Champernowne constant.

The more serious mistake is the conclusion drawn from the plot in Fig. 6 that  $C_{10}$  is not normal. The title “The construction of decimals normal in the scale of ten” of paper [2] itself claimed that  $C_{10}$  is normal. At the end of [2] Champernowne formulated theorems about normality of e.g. the number obtained by concatenating all consecutive composite natural numbers and conjectured that concatenation of prime numbers is also normal in base of ten. This last fact was later proved and generalized by A.H. Copeland and P. Erdős in the paper [5] (freely available at the BAMS website <http://projecteuclid.org/euclid.bams/1183509721>) to the general numbers  $0.a_1a_2a_3\dots$  which are normal provided the sequence  $a_n$  fulfills some growth condition. Presently the concatenation of “0.” with the base 10 representations of the prime numbers in order 0.2357111317192329... is called the Copeland–Erdős constant.

The conclusion by B. Kozarzewski about non normality of  $C_{10}$  is based on Fig. 6 showing frequency of digits 1, 4 and 8 in the first 2000000 digits of the Champernowne constant. In fact, Fig. 6 does not show the whole subtlety of the plot of digit frequency as there seems to be too large “coarse graining” used: the steps between two consecutive plotted points are so long that the finer structure is lost. In Fig. 1 we repro-

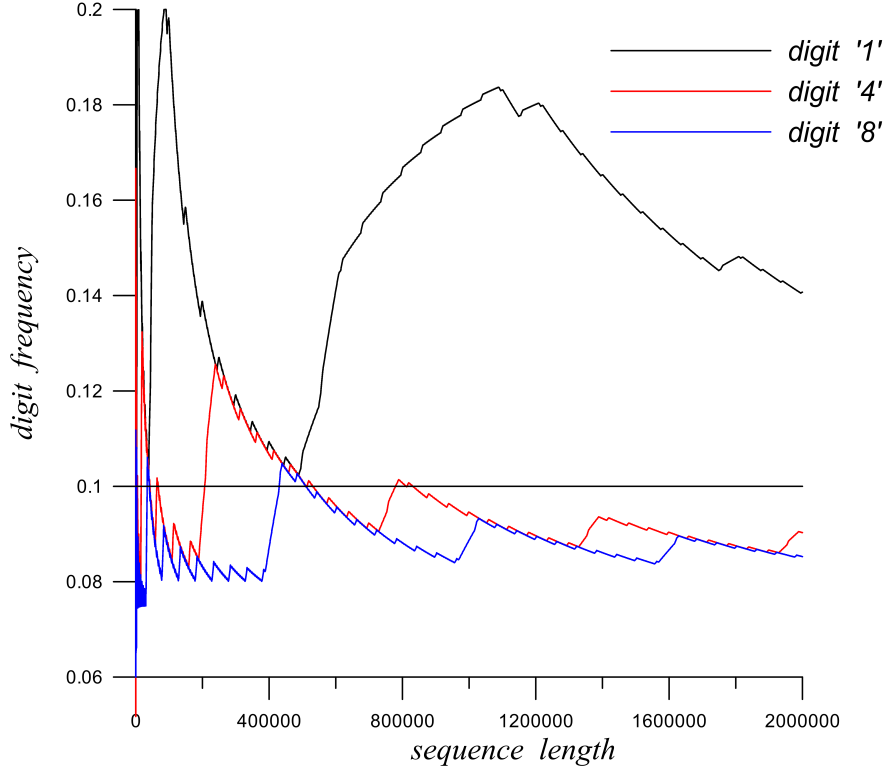


Fig. 1. The plot showing share of digits 1, 4 and 8 among the first 2000000 digits of  $C_{10}$ . There is a half million plotted points: digit frequency every 4-th position in  $C_{10}$  is plotted

duce the original plot from [1] this time with every 4-th point plotted, i.e. for each digit 1, 4, 8 there are 500000 points on the graphs. To see that the graphs of the digits frequency tend to 0.1 we present Fig.2 where the plots are made for all digits 0, 1, ..., 9 up to  $10^{11}$  position in the Champernowne constant. The characteristic oscillating pattern is caused by the mechanism creating  $C_{10}$ . We will denote the Champernowne number after concatenating the natural number  $n$  by  $C_{10}(n)$  and the running position in  $C_{10}$  will be denoted by  $\ell$ , while the position of the last digit of concatenated number  $n$  will be denoted by  $\ell(n)$ . Because successive digits of  $C_{10}$  come from successive natural numbers, there are local excesses of consecutive 1's, 2's, ..., 9's, resulting from runs of natural numbers beginning with 1, 2, ..., 9, respectively. For example: when five-digit natural numbers are concatenated first there are consecutive 10000 numbers beginning with 1: 10000, 10001, 10002, ..., 19999, 20000, 20001, ... Next there will be surplus of digits '2', after which the following digits will appear as the first digit of natural numbers. At the end of such a run of five-digit natural numbers the number 99999 appears and at the position of  $C_{10}$  marked by the last 9 of 99999 all digits 1, 2, ..., 9 will appear exactly the same number of times. There is a deficiency of digit '0' as no natural positive number has it as the first digit. There are 90 two-digit numbers in the base 10: 10, 11, ..., 99, next 900 three-digit numbers 100, 101, ..., 999 etc, in general there

are  $9 \times 10^{k-1}$   $k$ -digit natural numbers in the base 10. After concatenating all numbers from 1 to  $\underbrace{999 \dots 99}_{n \text{ nines}} (= 10^n - 1)$

the digits 1, 2, ..., 9 will appear exactly the same number of times (there is a dearth of zeroes). Each  $k$ -digits natural number increases the number of digits in  $C_{10}$  by  $k$ . Let us denote the number of digits in  $C_{10}$  after concatenating all numbers from 1 to  $10^n - 1$  by  $\ell(10^n - 1)$ . The value of  $\ell(10^n - 1)$  is given by the formula:

$$\begin{aligned} \ell(10^n - 1) &= \sum_{k=1}^n 9 \times k \times 10^{k-1} = \\ &= \frac{1}{9} (1 - 10^n + 9n \cdot 10^n) \end{aligned} \quad (1)$$

There are nine zeroes in the run of 90 two-digit natural numbers 10, 11, ..., 99, next in the run of 900 three-digit natural numbers 100, 101, ..., 999 the zero can appear only on the second and third positions, and as all digits appear on these second and third positions the same number of times the zero will appear  $2 \times 900/10 = 180$  times. In general, in the run  $9 \times 10^{n-1}$  of  $n$  digits natural numbers zero will appear  $9(n-1)10^{n-2}$  times. The number of zeroes  $n_0(n)$  in all numbers from 1 to  $999 \dots 99$  ( $n$  digits 9) hence is given by the formula

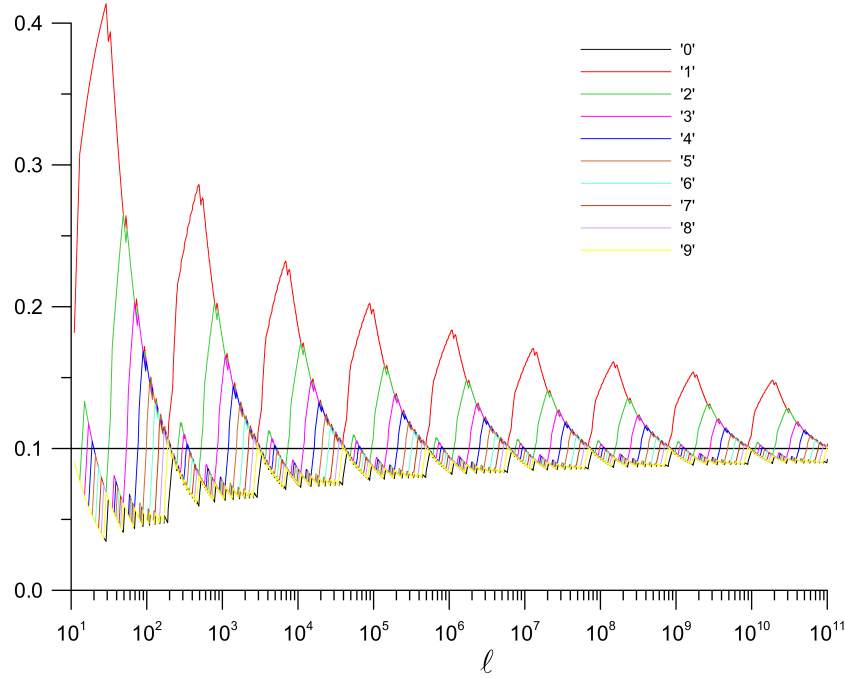


Fig. 2. The plot of the running frequency for digits  $0, 1, \dots, 9$  in  $C_{10}$  as a function of  $\ell$ . The data was collected at the points forming the geometrical progression  $10 \times (1.0002)^k$  and there are over 115000 of them. Obviously not all points are distinguishable on this graph: for resolution 600 dpi and the side of the  $x$ -axis say four inches only  $4 \cdot 600 = 2400$  points can be theoretically discerned; however, by collecting over 100000 data points we avoid a danger of dropping some abrupt change in the behavior of the functions for  $\ell = 11, \dots, 10^{11}$ . For all  $\ell$  the frequency of zero  $n_0(\ell)$  is below 0.1

$$\begin{aligned} n_0(10^n - 1) &= \sum_{k=2}^n 9(k-1)10^{k-2} = \\ &= \frac{1}{9} (1 - 10^n + 9n10^{n-1}) \end{aligned}$$

At the position  $\ell(10^n - 1)$  the number of digits  $1, 2, \dots, 9$  is the same and is given as  $1/9$  of the result of subtracting the number of zeroes  $n_0(10^n - 1)$  from  $\ell(10^n - 1)$ :

$$n_k(10^n - 1) = n10^{n-1} \quad k = 1, 2, \dots, 9$$

The limit  $n \rightarrow \infty$  of the frequencies  $f_k(10^n - 1) \equiv \frac{n_k(10^n - 1)}{\ell(10^n - 1)}$  is obviously 0.1.

Finally we would like to correct the definition of the algebraic number given in the left column on page 97. It should read "Some irrational numbers can be expressed as *roots* of a polynomial with rational coefficients".

## References

- [1] B. Kozarzewski, *A new method for symbolic sequences analysis. an application to long sequences*, CMST, **20**(3), 93-100 (2014).
- [2] D.G. Champernowne, *The construction of decimals normal in the scale of ten*, Journal of the London Mathematical Society **8**, 254-260 (1933).
- [3] M.M. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Matematico di Palermo **27**(1), 247-271 (1909).
- [4] K. Mahler, *Arithmetische Eigenschaften einer Klasse von Dezimalbrüchen*, Proc. Konin. Neder. Akad. Wet. Ser. A **40**, 421-428 (1937).
- [5] A.H. Copeland and P. Erdős, *Note on normal numbers*, Bull. Amer. Math. Soc. **52**, 857-860 (1946).



**Marek Wolf**, obtained PhD degree in physics (1982) and habilitation (1993) at the Wrocław University, where he was employed till 2011. Since 1984 he has been doing computer experiments in physics and mathematics. In 1991 and 1993 he was a research fellow at the Center for Polymer Studies at the Boston University. Since 2011 he is a member of the Faculty of Mathematics and Natural Sciences, College of Sciences at the Cardinal Stefan Wyszyński University in Warsaw. His hobby is photography.