PROCEEDINGS OF THE ROYAL SOCIETY A

FirstCite® e-publishing

Proc. R. Soc. A
doi:10.1098/rspa.2009.0126
Published online

The first-digit frequencies of prime numbers and Riemann zeta zeros

By Bartolo Luque and Lucas Lacasa*

Departamento de Matemática Aplicada y Estadística, ETSI Aeronáuticos, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Prime numbers seem to be distributed among the natural numbers with no law other than that of chance; however, their global distribution presents a quite remarkable smoothness. Such interplay between randomness and regularity has motivated scientists across the ages to search for local and global patterns in this distribution that could eventually shed light on the ultimate nature of primes. In this paper, we show that a generalization of the well-known first-digit Benford's law, which addresses the rate of appearance of a given leading digit d in datasets, describes with astonishing precision the statistical distribution of leading digits in the prime number sequence. Moreover, a reciprocal version of this pattern also takes place in the sequence of the non-trivial Riemann zeta zeros. We prove that the prime number theorem is, in the final analysis, responsible for these patterns.

Keywords: first significant digit; Benford's law; prime number; pattern; Riemann zeta function; counting function

1. Introduction

The individual location of prime numbers within the integers seems to be random; however, their global distribution exhibits a remarkable regularity (Zagier 1977). Certainly, this tension between local randomness and global order has led the distribution of primes to be, since antiquity, a fascinating problem for mathematicians (Dickson 2005) and, more recently, for physicists (Berry & Keating 1999; Kriecherbauer *et al.* 2001; Watkins, M. *Number theory & physics archive*, http://www.secamlocal.ex.ac.uk/people/staff/mrwatkin/zeta/physics.htm). The prime number theorem, which addresses the global smoothness of the counting function $\pi(n)$ providing the number of primes less or equal to integer n, was the first hint of such regularity (Tenenbaum & France 2000). Some other prime patterns have been advanced so far, from the visual Ulam spiral (Stein *et al.* 1964) to the arithmetic progression of primes (Green & Tao in press), while some others remain conjectures, such as the global gap distribution between primes or the twin primes distribution (Tenenbaum & France 2000), enhancing the mysterious interplay between apparent randomness and hidden regularity. There are indeed many open problems still to be solved, and the prime number distribution is yet to be understood (Guy 2004; Ribenboim 2004;

* Author for correspondence (lucas_lacasa@yahoo.es).

Received 6 March 2009
Accepted 23 March 2009

Caldwell, C. *The prime pages*, http://primes.utm.edu/). For instance, deep connections exist between the prime number sequence and the non-trivial zeros of the Riemann zeta function (Edwards 1974; Watkins, M. *Number theory & physics archive*, http://www.secamlocal.ex.ac.uk/people/staff/mrwatkin/zeta/physics.htm). The celebrated Riemann hypothesis, one of the most important open problems in mathematics, states that the non-trivial zeros of the complex-valued Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ (as a matter of fact, the meromorphic continuation of the function to the entire complex plane) are all complex numbers with real part 1/2, the location of these being intimately connected with the prime number distribution (Edwards 1964; Chernoff 2000).

Here, we address the statistics of the first significant or leading digit of both the sequences of primes and the sequence of the non-trivial Riemann zeta zeros. We show that while the first-digit distribution is asymptotically uniform in both sequences (that is to say, integers 1, ..., 9 tend to be equally likely as first digits in both sequences when we take into account the infinite amount of them), this asymptotic uniformity is reached in a very precise trend, namely by following a size-dependent generalized Benford's law (GBL), which constitutes an as yet unnoticed pattern in both sequences.

The rest of the paper is organized as follows. In §2, we introduce the most famous first-digit distribution: Benford's law. In §3, we introduce a generalization of Benford's law, and we show that both the sequences of the prime numbers and Riemann zeta zeros follow what we call a size-dependent GBL, introducing two unnoticed patterns of statistical regularity. In §4, we point out that the mean local density of both sequences is responsible for these latter patterns. We provide statistical arguments (statistical conformance between distributions) that support our claim. In §5, we provide some analytical arguments that confirm it. Specifically, making use of asymptotic expansion methods we prove that the prime number distribution is equivalent (within a margin of error) to a distribution that strictly follows a size-dependent GBL. At this point, we come up with new expressions for both the primes and the zeta zeros counting functions, precisely based on the pattern's structure previously found. In §6, we conclude and discuss possible applications.

2. Benford's law

The leading digit of a number represents its non-zero leftmost digit. For instance, the leading digits of the prime 7703 and the zeta zero 21.022... are 7 and 2, respectively. The most celebrated leading digit distribution is the so-called Benford's law (Hill 1996), after physicist Benford (1938), who empirically found that in many disparate natural datasets and mathematical sequences, the leading digit d was not uniformly distributed as might be expected, but instead had a biased probability as follows:

$$P(d) = \log_{10}(1 + 1/d), \tag{2.1}$$

where $d = 1, 2, \ldots, 9$. While this empirical law was indeed first discovered by astronomer Newcomb (1881), it is popularly known as Benford's law or, alternatively, as the law of anomalous numbers. Several disparate datasets such as stock prices, freezing points of chemical compounds or physical constants

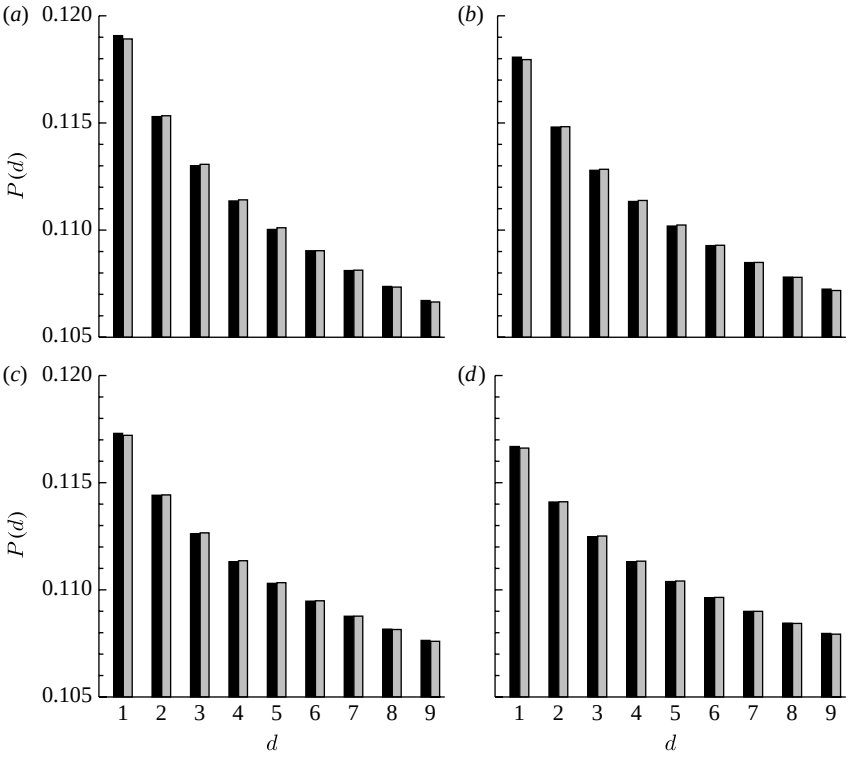

Figure 1. Leading digit histogram of the prime number sequence. Each plot represents, for the set of prime numbers in the interval $[1,N]$, the relative frequency of the leading digit d (black bars). Sample sizes are as follows: (*a*) 5761455 primes for $N=10^8$ ($\alpha=0.0583$); (*b*) 50847534 primes for $N=10^9$ ($\alpha=0.0513$); (*c*) 455052511 primes for $N=10^{10}$ ($\alpha=0.0458$); and (*d*) 4118054813 primes for $N=10^{11}$ ($\alpha=0.0414$). Grey bars represent the fit to a generalized Benford distribution (equation (3.1)) with a given exponent $\alpha(N)$.

exhibit this pattern at least empirically. While originally being only a curious pattern (Raimi 1976), practical implications began to emerge in the 1960s in the design of efficient computers (see, for instance, Knuth 1997). In recent years, goodness-of-fit tests against Benford's law have been used to detect possibly fraudulent data, by analysing the deviations of accounting data, corporation incomes, tax returns or scientific experimental data to theoretical Benford predictions (Nigrini 2000). Indeed, digit pattern analysis can produce valuable findings not revealed at first glance, as in the case of recent election results (Nigrini 2000; Mebane 2006).

Many mathematical sequences such as $(n^n)_{n\in\mathbb{N}}$ and $(n!)_{n\in\mathbb{N}}$ (Benford 1938), binomial arrays $\binom{n}{k}$ (Diaconis 1977), geometric sequences or sequences generated by recurrence relations (Raimi 1976; Miller & Takloo-Bighash 2006), to cite a few, have been proved to conform to Benford. Therefore, one may wonder if this is the case for the primes. In figure 1, we have plotted the leading digit d rate of appearance for the prime numbers placed in the interval $[1,N]$ (black bars), for different sizes N. Note that intervals $[1,N]$ have been chosen such that $N=10^D$, $D\in\mathbb{N}$ in order to assure an unbiased sample where all possible first digits are equiprobable *a priori* (see appendix A *a* for a discussion on natural densities and prime

numbers). Benford's law states that the first digit of a datum extracted at random is 1 with a frequency of 30.1 per cent, and is 9 only approximately 4.6 per cent. In figure 1, note that primes seem, however, to approximate to uniformity in their first digit. Indeed, the more we increase the interval under study, the more we approach uniformity (in the sense that all integers 1, …, 9 tend to be equally likely as a first digit). As a matter of fact, Diaconis (1977) proved that primes are not Benford distributed as long as their first significant digit is asymptotically uniformly distributed. A direct question arises: how does the prime sequence reach this uniform behaviour in the infinite limit? Is there any pattern on its trend towards uniformity, or, to the contrary, does the first-digit distribution lack any structure for finite sets?

3. Generalized Benford's law

Several mathematical insights regarding Benford's law have also been put forward so far (Pinkham 1961; Raimi 1976; Hill 1995*a*; Miller & Takloo-Bighash 2006), and Hill (1995*b*) proved a central limit-like theorem which states that random entries picked from random distributions form a sequence whose first-digit distribution tends towards Benford's law, explaining thereby its ubiquity. Practically, this law has for a long time been the only distribution that could explain the presence of skewed first-digit frequencies in generic datasets. Recently, Pietronero *et al.* (2001) proposed a generalization of Benford's law based on multiplicative processes (see also Nigrini & Miller 2007). It is well known that a stochastic process with probability density $1/x$ generates data that are Benford; therefore, series generated by power-law distributions $P(x) \sim x^{-\alpha}$, with $\alpha \neq 1$, would have a first-digit distribution that follows a so-called GBL,

$$P(d) = C \int_d^{d+1} x^{-\alpha}\, \mathrm{d}x = \frac{1}{10^{1-\alpha} - 1} [(d+1)^{1-\alpha} - d^{1-\alpha}], \tag{3.1}$$

where the prefactor is fixed for normalization to hold and α is the exponent of the original power-law distribution (observe that for $\alpha = 1$ the GBL reduces to the Benford law, while, for $\alpha = 0$, it reduces to the uniform distribution).

(*a*) *The first-digit frequencies of prime numbers*

Although Diaconis showed that the leading digit of primes distributes uniformly in the infinite limit, there exists a clear bias from uniformity for finite sets (figure 1). In this figure, we have also plotted (grey bars) the fitting to a GBL. Note that in each of the four intervals that we present, there is a particular value of exponent α for which an excellent agreement holds (see appendix A for fitting methods and statistical tests). More specifically, given an interval $[1,N]$, there exists a particular value $\alpha(N)$ for which a GBL fits with extremely good accuracy the first-digit distribution of the primes appearing in that interval. Observe at this point that the functional dependency of α is only in the interval's upper bound; once this bound is fixed, α is constant in that interval. Interestingly, the value of the fitting parameter α decreases as the interval's upper bound N, hence the number of primes, increases. In figure 2*a*, we have plotted this size dependence, showing that a functional relation between α and N seems to take place,

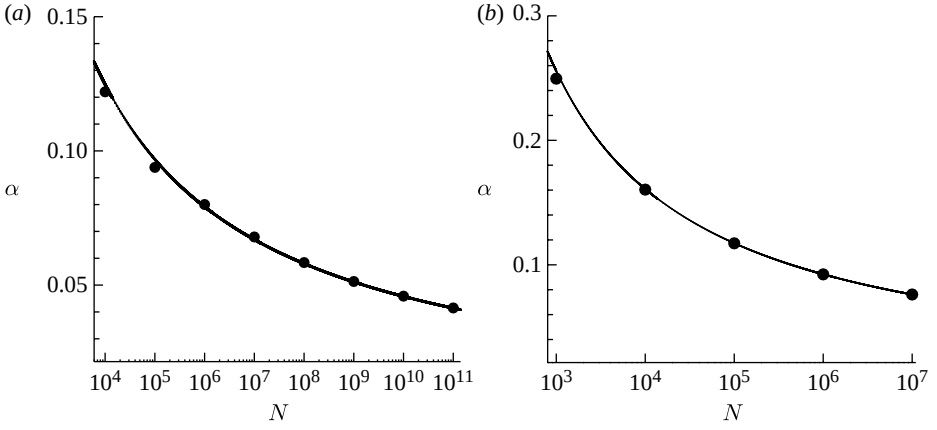

Figure 2. Size-dependent parameter $\alpha(N)$. (*a*) Circles represent the exponent $\alpha(N)$ for which the first significant digit of the prime number sequence fits a generalized Benford law in the interval $[1,N]$. The black line corresponds to the fit, using a least-squares method, $\alpha(N) = 1/(\log N - 1.10)$. (*b*) The same analysis as for (*a*), but for the non-trivial Riemann zeta zeros sequence. The best fit is $\alpha(N) = 1/(\log N - 2.92)$.

$$\alpha(N) = \frac{1}{\log N - a}, \tag{3.2}$$

where $a = 1.1 \pm 0.1$ is the best fit. Note that $\lim_{N\to\infty} \alpha(N) = 0$, and this size-dependent GBL, reduces asymptotically to the uniform distribution, which is consistent with previous theory (Diaconis 1977). Despite the local randomness of the prime number sequence, it seems that its first-digit distribution converges smoothly to uniformity in a very precise trend: as a GBL with a size-dependent exponent $\alpha(N)$.

At this point and as in the case of Benford's law (Hill 1995*b*), an extension of the GBL to include not only the first significative digit but also the first k significative ones can be performed. Given a number n, we can consider its k first significative digits d_1, d_2, …, d_k through its decimal representation: $D = \sum_{i=1}^{k} d_i 10^{k-i}$, where $d_1 \in \{1, \ldots, 9\}$ and $d_i \in \{0, 1, \ldots, 9\}$ for $i \geq 2$. Hence, the *extended* GBL providing the probability of starting with number D is

$$P(d_1, d_2, \ldots, d_k) = P(D) = \frac{1}{(10^k)^{1-\alpha} - 10^{k-1}} [(D+1)^{1-\alpha} - D^{1-\alpha}]. \tag{3.3}$$

Figure 3 represents the fit of the 4118054813 primes appearing in the interval $[1,10^{11}]$ to an *extended* GBL for k=2,3,4 and 5: interestingly, the pattern still holds.

(*b*) *The 'mirror' pattern in the Riemann zeta zeros sequence*

Since prime numbers are strongly related to the non-trivial Riemann zeta zeros, one may wonder if a similar pattern holds in this latter sequence (zeros sequence from now on). This sequence is composed of the imaginary part of the non-trivial zeros of $\zeta(s)$ (actually only those with a positive imaginary part are taken into account for reasons of symmetry, since the zeros are symmetrically

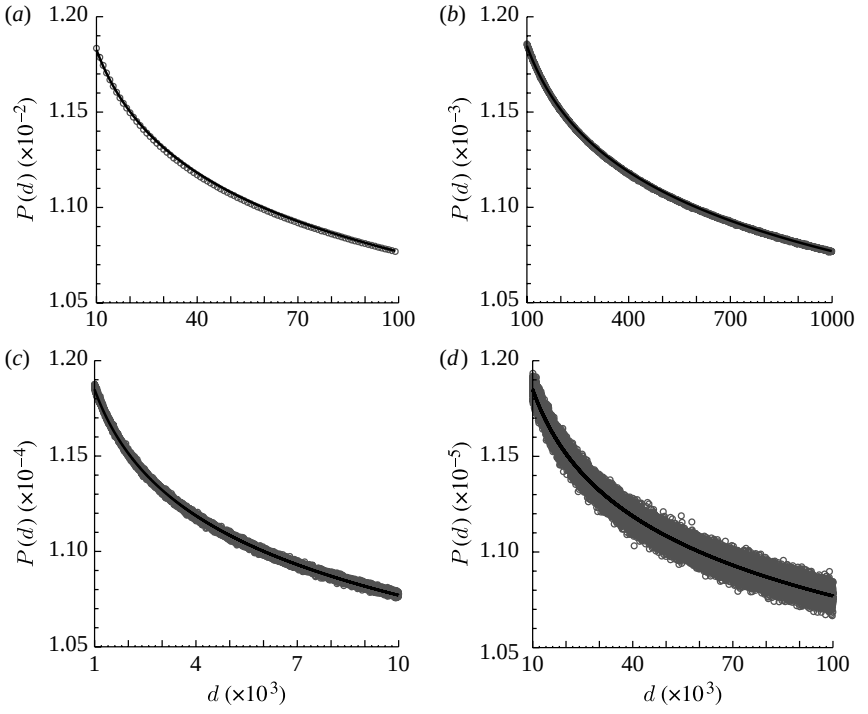

Figure 3. Extension of GBL to the k first significant digits. We represent the fit of an extended GBL following equation (3.3) (black line) to (*a*) the first two significant digits' relative frequencies, (*b*) first three significant digits' relative frequencies, (*c*) first four significant digits' relative frequencies and (*d*) first five significant digits' relative frequencies of the 4118054813 primes appearing in the interval $[1,10^{11}]$ (circles). Observe that the fit has been made with $\alpha(N=10^{11})=0.0414$ in every case (equation (3.2)).

distributed about the central point). This sequence is not Benford distributed according to a theorem by Rademacher and Hlawka (Hlawka 1984) which proves that it is *asymptotically* uniform. Nevertheless, will it follow a size-dependent GBL as in the case of the primes?

In figure 4, we have plotted, in the interval $[1,N]$ and for different values of N, the relative frequencies of leading digit d in the zeros sequence (black bars), and in grey bars a fit to a GBL with density x^{α}, i.e.

$$P(d)=C\int_{d}^{d+1}x^{\alpha}\,\mathrm{d}x=\frac{1}{10^{1+\alpha}-1}[(d+1)^{1+\alpha}-d^{1+\alpha}] \tag{3.4}$$

(this reciprocity is clarified later in the text). Note that a very good agreement holds again for particular size-dependent values of α, and the same functional relation as equation (3.2) holds, with $a=2.92\pm0.05$ as the best fit. As in the case of the primes, this size-dependent GBL tends to uniformity for $N\rightarrow\infty$, as it should (Hlawka 1984). Moreover, the *extended* version of equation (3.4) for the k first significative digits is

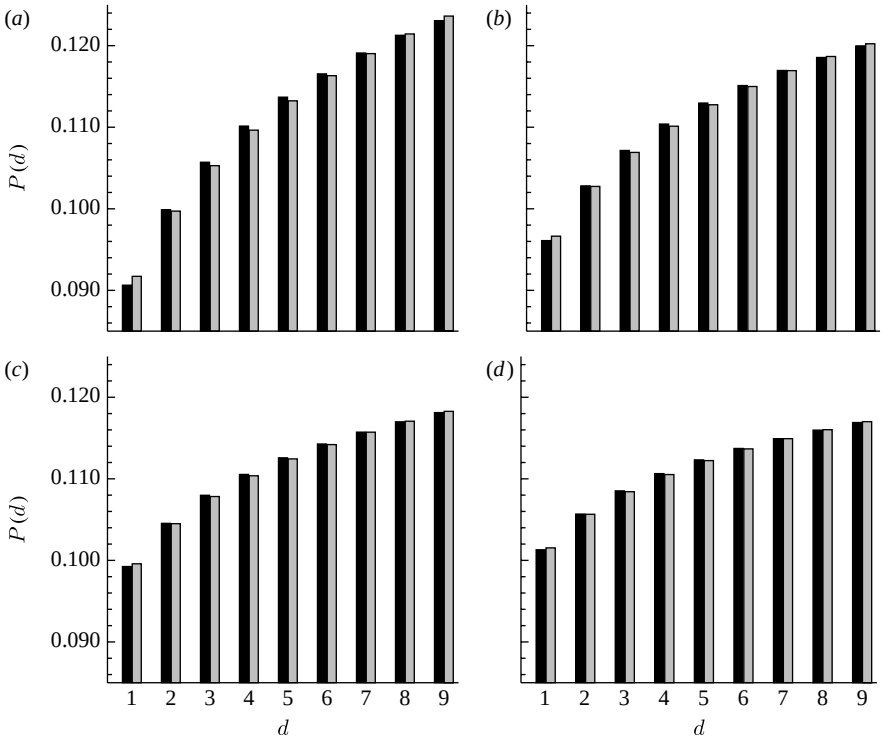

Figure 4. Leading digit histogram of the non-trivial Riemann zeta zeros sequence. Each plot represents, for the sequence of the Riemann zeta zeros in the interval $[1,N]$, the observed relative frequency of leading digit d (black bars). Sample sizes are: (*a*) 10142 zeros for $N=10^4$ ($\alpha=0.1603$), (*b*) 138069 zeros for $N=10^5$ ($\alpha=0.1172$), (*c*) 1747146 zeros for $N=10^6$ ($\alpha=0.0923$) and (*d*) 21136126 zeros for $N=10^7$ ($\alpha=0.0761$). Grey bars represent the fit to a GBL following equation (3.4) with a given exponent $\alpha(N)$.

$$P(d_1, d_2, \ldots, d_k) = P(D) = \frac{1}{(10^k)^{1+\alpha} - 10^{k-1}}[(D+1)^{1+\alpha} - D^{1+\alpha}]. \tag{3.5}$$

As can be seen in figure 5, the pattern also holds in this case.

4. Statistical conformance of prime number distribution to GBL

Why do these two sequences exhibit this unexpected pattern in the leading digit distribution? What is causing it to take place? While the prime number distribution is deterministic in the sense that precise rules determine whether an integer is prime or not, its apparent local randomness has suggested several stochastic interpretations. In particular, Cramér (1935, see also Tenenbaum & France 2000) defined the following model: assume that we have a sequence of urns $U(n)$, where $n=1, 2, \ldots$, and put black and white balls in each urn such that the probability of drawing a white ball in the k^{th}-urn goes as $1/\log k$. Then, in order to generate a sequence of pseudo-random prime numbers, we need only to draw a ball from each urn: if the drawing from the k^{th}-urn is white, then k will

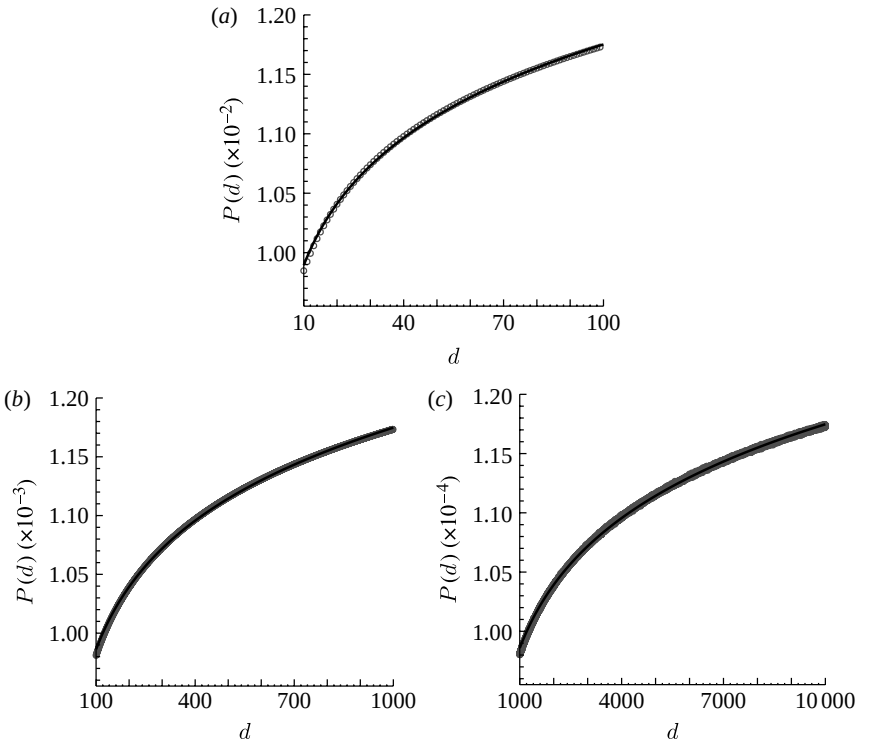

Figure 5. Extension of GBL to the k first significant digits. We represent the fit of an extended GBL following equation (3.5) (black line) to (*a*) the first two significant digits' relative frequencies, (*b*) first three significant digits' relative frequencies and (*c*) first four significant digits' relative frequencies of the 21136126 zeros appearing in the interval $[1,10^7]$ (circles). Observe that in every case, the fit has been performed with $\alpha(N=10^7)=0.0761$.

be labelled as a pseudo-random prime. The prime number sequence can indeed be understood as a concrete realization of this stochastic process, where the chance of a given integer x to be prime is $1/\log x$.

We have repeated all the statistical tests within the stochastic Cramér model, and have found that a statistical sample of pseudo-random prime numbers in $[1,10^{11}]$ is also GBL distributed and reproduces all the statistical analyses previously found in the actual primes (see appendix A for an in-depth analysis). This result strongly suggests that a density $1/\log x$, which is nothing but the mean local prime density by virtue of the prime number theorem, is likely to be responsible for the GBL pattern. In the following we provide further statistical and analytical arguments that support this fact.

Recently, it has been shown that disparate distributions such as the lognormal, the Weibull or the exponential distribution can generate standard Benford behaviour (Leemis *et al.* 2000) for particular values of their parameters. In this sense, a similar phenomenon could be taking place with GBL: can different distributions generate GBL behaviour? One should thus switch the emphasis from the examination of datasets that obey GBL to *probability distributions* that do so, other than power laws.

Table 1. χ^2 goodness-of-fit test c of the conformance between cumulative distributions of primes ($\pi(x)/\pi(N)$ and $\mathrm{Li}(x)/\mathrm{Li}(N)$) and a GBL with exponent $\alpha(N)$ (equation (3.2)) in the interval $[1,N]$. (The null hypothesis that prime number distribution obeys GBL cannot be rejected.)

N	c for $\pi(x)/\pi(N)$	c for $\mathrm{Li}(x)/\mathrm{Li}(N)$
10^3	0.59×10^{-2}	0.58×10^{-2}
10^4	0.86×10^{-3}	0.57×10^{-3}
10^5	0.12×10^{-3}	0.13×10^{-3}
10^6	0.57×10^{-4}	0.61×10^{-4}
10^7	0.32×10^{-4}	0.33×10^{-4}
10^8	0.17×10^{-4}	0.17×10^{-4}

(*a*) χ^2-*test for conformance between distributions*

The prime counting function $\pi(N)$ provides the number of primes in the interval $[1,N]$ (Tenenbaum & France 2000) and, up to normalization, stands as the cumulative distribution function of primes. While $\pi(N)$ is a stepped function, a nice asymptotic approximation is the offset logarithmic integral

$$\pi(N) \sim \int_2^N \frac{1}{\log x}\mathrm{d}x = \mathrm{Li}(N) \tag{4.1}$$

(one of the formulations of the Riemann hypothesis actually states that $|\mathrm{Li}(n)-\pi(n)| < c\sqrt{n}\log n$, for some constant c; Edwards 1974). We can interpret $1/\log x$ as an average prime density and the lower bound of the integral is set to be 2 for singularity reasons. Following Leemis *et al.* (2000), we can calculate a χ^2 goodness-of-fit test of the conformance between the first-digit distribution generated by $\mathrm{Li}(N)$ and a GBL with exponent $\alpha(N)$. The test statistic is in this case

$$c = \sum_{d=1}^{9} \frac{[\Pr(Y=d)-\Pr(X=d)]^2}{\Pr(X=d)}, \tag{4.2}$$

where $\Pr(X)$ is the first-digit probability (equation (3.1)) for a GBL associated to a probability distribution with exponent $\alpha(N)$, and $\Pr(Y)$ is the tested probability. In table 1, we have computed, fixed the interval $[1,N]$, the χ^2-statistic c for two different scenarios, namely the normalized logarithmic integral $\mathrm{Li}(n)/\mathrm{Li}(n)$ and the normalized prime counting function $\pi(n)/\pi(N)$, with $n\in[1,N]$. In both cases, there is a remarkable good agreement and we cannot reject the hypothesis that primes are size-dependent GBL.

(*b*) *Conditions for conformance to GBL*

Hill (1995*b*) wondered about which common distributions (or mixtures thereof) satisfy Benford's law. Leemis *et al.* (2000) tackled this problem and quantified the agreement to Benford's law of several standard distributions. They concluded that the ubiquity of Benford behaviour could be related to the fact that many distributions follow Benford's law for particular values of

their parameters. Here, following the philosophy of that work (Leemis *et al.* 2000), we develop a mathematical framework that provides conditions for conformance to a GBL.

The probability density function of a discrete GB random variable Y is

$$f_Y(y) = \Pr(Y = y) = \frac{1}{10^{1-\alpha} - 1}[(y+1)^{1-\alpha} - y^{1-\alpha}], \quad y = 1, 2, \ldots, 9. \tag{4.3}$$

The associated cumulative distribution function is therefore

$$F_Y(y) = \Pr(Y \leq y) = \frac{1}{10^{1-\alpha} - 1}[(y+1)^{1-\alpha} - 1], \quad y = 1, 2, \ldots, 9. \tag{4.4}$$

How can we prove that a random variable T extracted from a probability density $f_T(t) = \Pr(t)$ has an associated (discrete) random variable Y that follows equation (4.3)? We can readily find a relation between both random variables. Suppose, without loss of generality, that the random variable T is defined in the interval $[1,10^{D+1})$. Let the discrete random variable D fulfil

$$10^D \leq T < 10^{D+1}. \tag{4.5}$$

This definition allows us to express the first significative digit Y in terms of D and T,

$$Y = \lfloor T \cdot 10^{-D} \rfloor, \tag{4.6}$$

where from now on the floor brackets stand for the integer part function. Now, let U be a random variable uniformly distributed in (0,1), $U \sim U(0,1)$. Then, inverting the cumulative distribution function (4.4), we obtain

$$Y = \left\lfloor [(10^{1-\alpha} - 1) \cdot U + 1]^{1/(1-\alpha)} \right\rfloor. \tag{4.7}$$

This latter relation is useful to generate a discrete GB random variable Y from a uniformly distributed one $U(0,1)$. Note also that for $\alpha = 0$, we have $Y = \lfloor 9 \cdot U + 1 \rfloor$ that is a first-digit distribution which is uniform $\Pr(Y = y) = 1/9, \quad y = 1, 2, \ldots, 9$, as expected. Hence, every discrete random variable Y that distributes as a GB should fulfil equation (4.7), and, consequently, if a random variable T has an associated random variable Y, the following identity should hold:

$$\lfloor T \cdot 10^{-D} \rfloor = \left\lfloor [(10^{1-\alpha} - 1) \cdot U + 1]^{1/(1-\alpha)} \right\rfloor \tag{4.8}$$

and then,

$$Z = \frac{(T10^{-D})^{1-\alpha} - 1}{10^{1-\alpha} - 1} \sim U(0,1). \tag{4.9}$$

In other words, in order for the random variable T to generate a GB, the random variable Z defined in the preceding transformation should distribute as $U(0,1)$. The cumulative distribution function of Z is thus given by

$$F_Z(z) = \sum_{d=0}^{n} \left\{ \Pr(10^d \leq T < 10^{d+1}) \cdot \Pr\left(\frac{(T10^{-D})^{1-\alpha} - 1}{10^{1-\alpha} - 1} \leq z \,|\, 10^d \leq T < 10^{d+1} \right) \right\} = z, \tag{4.10}$$

that in terms of the cumulative distribution function of T becomes

$$\sum_{d=0}^{n} \{F_T(v10^d) - F_T(10^d)\} = z, \tag{4.11}$$

where $v \equiv [(10^{1-\alpha} - 1)z + 1]^{1/(1-\alpha)}$.

We may consider now the power-law density $x^{-\alpha}$ proposed by Pietronero *et al.* (2001) in order to show that this distribution exactly generates generalized Benford behaviour,

$$f_T(t) = \Pr(t) = \frac{1-\alpha}{10^{(D+1)(1-\alpha)}-1}\, t^{-\alpha}, \quad t \in [1, 10^{D+1}). \tag{4.12}$$

Its cumulative distribution function will be

$$F_T(t) = \frac{t^{1-\alpha}-1}{10^{(D+1)(1-\alpha)}-1} \tag{4.13}$$

and thereby equation (4.11) reduces to

$$\sum_{d=0}^{D}\{F_T(v10^d) - F_T(10^d)\} = \frac{z(10^{1-\alpha}-1)}{10^{(D+1)(1-\alpha)}-1}\sum_{d=0}^{D}(10^{1-\alpha})^d = z, \tag{4.14}$$

as expected.

(*c*) *GBL holds for prime number distribution*

While the preceding development is in itself interesting in order to check for the conformance of several distributions to GBL, we will restrict our analysis to the cumulative distribution function of the prime number conveniently normalized in the interval [1,10^D],

$$F_T(t) = \frac{\pi(t)}{\pi(10^{D+1})}, \quad t \in [1, 10^{D+1}). \tag{4.15}$$

Note that previous numerical analysis showed that

$$\alpha(10^{D+1}) = \frac{1}{\ln(10^{D+1}) - a}, \tag{4.16}$$

where $a \simeq 1.1$. Since $\pi(t)$ is a stepped function that does not possess a closed form, the relation (4.11) cannot be analytically checked. However, a numerical exploration can indicate into which extent primes are conformal with GBL. Relation (4.11) reduces in this case to check if

$$\sum_{d=0}^{D}\{\pi(v\cdot 10^d) - \pi(10^d)\} \approx \pi(10^{D+1})z, \tag{4.17}$$

where $v \equiv [(10^{1-\alpha(10^{D+1})}-1)z+1]^{1/(1-\alpha(10^{D+1}))}$ and $z \in [0,1]$. First, this latter relation is trivially fulfilled for the extremal values $z=0$ and 1. For other values $z \in (0,1)$, we have numerically tested this equation for different values of D, and have found that it is satisfied with negligible error (we have performed a scatterplot of equation (4.17) and have found a correlation coefficient $r=1.0$).

The same numerical analysis has been performed for logarithmic integral Li. In this case, the relation that must be fulfilled is

$$\sum_{d=0}^{D}\{\mathrm{Li}(v\cdot 10^d) - \mathrm{Li}(10^d)\} \approx \mathrm{Li}(10^{D+1})z \tag{4.18}$$

and is indeed satisfied with similar remarkable results provided that we fix $\mathrm{Li}(1) \equiv 0$ for singularity reasons.

5. Counting functions for prime numbers and zeta zeros

Hitherto, we have provided statistical arguments which indicate that other distributions than $x^{-\alpha}$ such as $1/\log x$ can generate GBL behaviour. In the following we provide analytical arguments that support this fact.

(a) The primes counting function $L(N)$

Suppose that a given sequence has a power-law-like density $x^{-\alpha}$ (and whose first significative digits are consequently GBL). One can derive from this latter density a counting function $L(N)$ that provides the number of elements of that sequence appearing in the interval $[1,N]$. A first option is to assume a local density of the shape $x^{-\alpha(x)}$, such that $L(N) \sim \int_2^N x^{-\alpha(x)}\,\mathrm{d}x$. Note that this option implicitly assumes that α varies smoothly in $[1,N]$, which is not the case in the light of the numerical relation (3.2), which implies that the functional dependency of α is only with respect to the upper bound value of the interval. Indeed, $x^{-\alpha(x)}$ is not a good approximation to $1/\ln x$ for any given interval. This drawback can be overcome defining $L(N)$ as follows:

$$L(N) = e\alpha(N)\int_2^N x^{-\alpha(N)}\,\mathrm{d}x, \tag{5.1}$$

where the prefactor is fixed for $L(N)$ to fulfil the prime number theorem and, consequently,

$$\lim_{N\to\infty}\frac{L(N)}{N/\log N} = 1 \tag{5.2}$$

(see table 2 for a numerical exploration of this new approximation to $\pi(N)$). Observe that what we are claiming is that the fixed interval $[1,N]$, $x^{-\alpha(N)}$ acts as a good approximation to the primes mean local density $1/\ln x$ in that interval. In order to prove it, let us compare the counting functions derived from both densities. First, $\mathrm{Li}(N) = \int_2^N (1/\ln x)\mathrm{d}x$ possesses the following asymptotic expansion:

$$\mathrm{Li}(N) = \frac{N}{\log N}\left\{1 + \frac{1}{\log N} + \frac{2}{\log^2 N} + O\left(\frac{1}{\log^3 N}\right)\right\}. \tag{5.3}$$

On the other hand, we can asymptotically expand $L(N)$ as it follows:

$$\begin{aligned} L(N) &= \frac{\alpha(N)e}{1-\alpha(N)}N^{1-\alpha(N)} \\ &= \frac{N}{\log N-(a+1)}\cdot\exp\left(\frac{-a}{\log N-a}\right) \\ &= \frac{N}{\log N}\left\{1+\frac{a+1}{\log N}+\frac{(a+1)^2}{\log^2 N}+O\left(\frac{1}{\log^3 N}\right)\right\}\cdot \\ &\quad \left\{1-\frac{a}{\log N-a}+\frac{a^2}{2(\log N-a)^2}+O\left(\frac{1}{(\log N-a)^3}\right)\right\} \\ &= \frac{N}{\log N}\left\{1+\frac{1}{\log N}+\frac{1+a-a^2/2}{\log^2 N}+O\left(\frac{1}{\log^3 N}\right)\right\}. \end{aligned} \tag{5.4}$$

Table 2. Up to integer N, values of the prime counting function $\pi(N)$, the approximation given by the logarithmic integral $\mathrm{Li}(N)$, $N/\log N$, the counting function $L(N)$ defined in equation (5.1) and the ratio $L(N)/\pi(N)$.

N	$\pi(N)$	$\mathrm{Li}(N)$	$N/\log N$	$L(N)$	$L(N)/\pi(N)$
10^2	25	30	22	29	0.85533
10^3	168	178	145	172	0.97595
10^4	1229	1246	1086	1228	1.00081
10^5	9592	9630	8686	9558	1.00352
10^6	78492	78628	72382	78280	1.00278
10^7	664579	664918	620421	662958	1.00244
10^8	5761455	5762209	5428681	5749998	1.00199
10^9	50847534	50849235	48254942	50767815	1.00157
10^{10}	455052511	455055615	434294482	454484882	1.00125
10^{20}	2220819602560918840				1.00027

Comparing equations (5.3) and (5.4), we conclude that $\mathrm{Li}(N)$ and $L(N)$ are compatible cumulative distributions within an error

$$E(N) = \frac{N}{\log N}\left\{\frac{2}{\log^2 N} - \frac{1+a-a^2/2}{\log^2 N} + O\left(\frac{1}{\log^3 N}\right)\right\}, \tag{5.5}$$

which is indeed minimum for $a=1$, consistent with our previous numerical results obtained for the fitting value of a (equation (3.2)). Hence, within that error, we can conclude that primes obey a GBL with $\alpha(N)$ following equation (3.2): primes follow a size-dependent GBL.

(*b*) *The zeta zeros counting function S(N)*

What about the Riemann zeros? Von Mangoldt proved (Edwards 1974) that, on average, the number of non-trivial zeros $R(N)$ up to N (zeros counting function) is

$$R(N) = \frac{N}{2\pi}\log\left(\frac{N}{2\pi}\right) - \frac{N}{2\pi} + O(\log N). \tag{5.6}$$

$R(N)$ is nothing but the cumulative distribution of the zeros (up to normalization), which satisfies

$$R(N) \approx \frac{1}{2\pi}\int_2^N \log\left(\frac{x}{2\pi}\right)\mathrm{d}x. \tag{5.7}$$

The non-trivial Riemann zeros average density is thus $\log(x/2\pi)$, which is essentially the reciprocal of the prime numbers mean local density (see equation (4.1)). One can thus straightforwardly deduce a power-law approximation $S(N)$ to the cumulative distribution $R(N)$ of the non-trivial zeros similar to equation (5.1),

$$S(N) \sim \frac{1}{2\pi e\alpha(N/2\pi)}\int_2^N \left(\frac{x}{2\pi}\right)^{\alpha(N/2\pi)}\mathrm{d}x. \tag{5.8}$$

We conclude that zeros are also GBL for $\alpha(N)$ satisfying the following change of scale:

$$\alpha(N/2\pi) = \frac{1}{\log(N/2\pi) - a} = \frac{1}{\log N - (\log(2\pi) + a)}. \tag{5.9}$$

Hence, since $a \simeq 1.1$ (equation (5.5)), one should expect the following value for the constant a associated to the zeros sequence: $\log(2\pi) + 1.1 \approx 2.93$, which is in good agreement with our previous numerical analysis.

6. Discussion

To conclude, we have unveiled a statistical pattern in the sequences of the prime numbers and the non-trivial Riemann zeta zeros that has surprisingly gone unnoticed until now. According to several statistical and analytical arguments, we can conclude that, for a fixed interval $[1,N]$, we can approximate the mean local density of both sequences to a power-law distribution with good accuracy, and this is indeed responsible for these patterns. Along with this finding, some relations concerning the statistical conformance of any given distribution to the generalized Benford law have also been derived.

Several applications and future work can be depicted: first, since the Riemann zeros seem to have the same statistical properties as the eigenvalues of a concrete type of random matrices called the Gaussian unitary ensemble (Berry & Keating 1999; Bogomolny 2007), the relation between GBL and random matrix theory should be investigated in depth (Miller & Kontorovich 2005). Second, this finding may also apply to several other sequences that, while not being strictly Benford distributed, can be GBL, and, in this sense, much work recently developed for Benford distributions (Hürlimann 2006) could be readily generalized. Finally, it has not escaped our notice that several applications recently advanced in the context of Benford's law, such as fraud detection or stock market analysis (Nigrini 2000), could eventually be generalized to the wider context of GBL formalism. This generalization also extends to stochastic sieve theory (Hawkins 1957), dynamical systems that follow Benford's law (Berger *et al.* 2005; Miller & Takloo-Bighash 2006) and their relation to stochastic multiplicative processes (Manrubia & Zanette 1999).

We thank I. Parra for helpful suggestions and K. McCourt, O. Miramontes, J. Bascompte, D. H. Zanette and S. C. Manrubia for their comments. This work was supported by grant FIS2006-08607 from the Spanish Ministry of Science.

Appendix A. Statistical methods and technical digressions

(a) *How to pick an integer at random?*

(i) *Visualizing the generalized Benford law pattern in prime numbers as a biased random walk*

In order for the pattern already captured in figure 1 to become more evident, we have built the following two-dimensional random walk:

$$x(t+1) = x(t) + \xi_x, \qquad y(t+1) = y(t) + \xi_y, \tag{A 1}$$

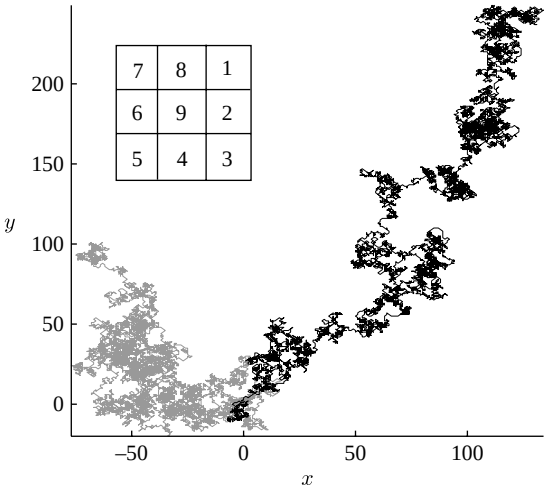

Figure 6. Random walks. Grey: two-dimensional random walk in which, at each step, we pick at random a natural from $[1,10^6]$ and move forward depending on the value of its first significative digit following the rules depicted in the inner table. The behaviour approximates an uncorrelated Brownian motion: integers' first digit is uniformly distributed. Black: the same random walk but picking at random primes in $[1,10^6]$; in this case, the random walk is clearly biased.

where x and y are Cartesian variables with $x(0) = y(0) = 0$, and both ξ_x and ξ_y are discrete random variables that have values $\in\{0, -1, 1\}$ depending on the first digit d of the numbers randomly chosen at each time step, according to the rules depicted in figure 6. Thereby, in each iteration, we peak at random a positive integer (grey random walk) or a prime (black random walk) from the interval $[1,10^6]$, and depending on its first significative digit d, the random walker moves accordingly (for instance, if we peak prime 13, we have $d=1$ and the random walker rules provide $\xi_x=1$ and $\xi_y=1$: the random walker moves up-right). We have plotted the results of this two-dimensional random walk in figure 6 for random picking of integers (grey random walk) and random picking of primes (black random walk). Note that while the grey random walk seems to be a typical uncorrelated Brownian motion (enhancing the fact that the first-digit distribution of the integers is uniformly distributed), the black random walk is clearly biased: this is indeed a visual characterization of the pattern. Observe that if the interval in which we randomly peak either the integers or the primes were not of the shape $[1,10^D]$, there would be a systematic bias present in the pool and, consequently, both integer and prime random walks would be biased; it is therefore necessary to define the intervals under study in that way.

(ii) *Natural density*

If primes were, for instance, Benford distributed, one should expect that if we pick a prime at random, this one should start with the number 1 around 30 per cent of the time. But what does the sentence 'pick a prime at random' mean? Note that in the previous experiment (the two-dimensional biased random walk), we have drawn numbers, whether integers or primes, at random from the pool $[1,10^6]$. Throughout this paper, the intervals $[1,N]$ have been chosen so that $N=10^D$, $D \in \mathbb{N}$. This choice is not arbitrary, but very much to the contrary,

it relies on the fact that whenever studying infinite integer sequences, the results strongly depend on the interval under study. For instance, everyone would agree that intuitively the set of positive integers $\mathbb{N}$ is an infinite sequence whose first digit is uniformly distributed: there exist as many naturals starting with 1 as naturals starting with 9. However, there exist subtle difficulties at this point arising from the fact that the first-digit natural density is not well defined. Since there exist infinite integers in $\mathbb{N}$ and, consequently, it is not straightforward to quantify the quote 'pick an integer at random' in a way which satisfies the laws of probability, in order to check if integers have a uniform distributed first significant digit, we have to consider finite intervals $[1,N]$. Hereafter, note that uniformity *a priori* is only respected when $N=10^D$. For instance, if we choose the interval to be [1,2000] and we randomly draw a number, this one will start with 1 with rather large probability, as there are obviously more numbers starting by one in that interval. If we increase the interval to say [1,3000], then the probability of drawing a number starting with 1 or 2 will be larger than the probability of any other. We can easily come to the conclusion that the first-digit density will oscillate repeatedly by decades as N increases without reaching convergence, and it is thereby said that the set of positive integers with leading digit d ($d=1, 2, \ldots, 9$) does not possess a natural density among the integers. Note that the same phenomenon is likely to take place for the primes (see Chris Caldwell's *The prime pages* for an introductory discussion of natural density and Benford's law for prime numbers, http://primes.utm.edu/ and references therein).

In order to overcome this subtle point, one can (i) choose intervals of the shape $[1,10^D]$, where every leading digit has equal probability *a priori* of being picked. According to this situation, positive integers $\mathbb{N}$ have a uniform first-digit distribution, and, in this sense, Diaconis (1977) showed that primes do not obey Benford's law as their first-digit distribution is asymptotically uniform. Or (ii) use average and summability methods such as the Cesaro or the logarithm matrix method $\mathfrak{l}$ (Raimi 1976) in order to define a proper first-digit density that holds in the infinite limit. Some authors have shown that, in this case, both the primes and the integers are said to be *weak* Benford sequences (Flehinger 1966; Withney 1972; Raimi 1976).

As we are dealing with finite subsets and in order to check if a pattern *really* takes place for the primes, in this work, we have chosen intervals of the shape $[1,10^D]$ to assure that samples are unbiased and that all first digits are equiprobable *a priori*.

(*b*) *Statistical methods*

(i) *Method of moments*

In order to estimate the best fit between a GBL with parameter α and a dataset, we have employed the method of moments. If GBL fits the empirical data, then both distributions have the same first moments, and the following relation holds:

$$\sum_{d=1}^{9} dP(d) = \sum_{d=1}^{9} dP^e(d), \tag{A 2}$$

Table 3. Table gathering the values of the following statistics: χ^2, maximum absolute deviation (m), mean absolute deviation (MAD) and correlation coefficient (r) between the observed first significant digit frequency of the set of M primes in the interval $[1,N]$ and the expected generalized Benford distribution (equation (3.1) with an exponent $\alpha(N)$ given by equation (3.2) with $a=1.1$). (While the χ^2-test rejects the hypothesis for very large samples due to its size sensitivity, every other test cannot reject it, enhancing the goodness of fit between the data and the generalized Benford distribution.)

N	M=no. of primes	χ^2	m	MAD	r
10^4	1229	0.45	0.32×10^{-2}	0.19×10^{-2}	0.96965
10^5	9592	0.62	0.21×10^{-2}	0.65×10^{-3}	0.99053
10^6	78498	0.61	0.50×10^{-3}	0.26×10^{-3}	0.99826
10^7	664579	0.77	0.17×10^{-3}	0.11×10^{-3}	0.99964
10^8	5761455	2.2	0.15×10^{-3}	0.56×10^{-4}	0.99984
10^9	50847534	11.0	0.11×10^{-3}	0.42×10^{-4}	0.99988
10^{10}	455052511	61.2	0.90×10^{-4}	0.33×10^{-4}	0.99991
10^{11}	4118054813	358.5	0.74×10^{-4}	0.27×10^{-4}	0.99993

where $P(d)$ and $P^e(d)$ are the observed normalized frequencies and GB expected probabilities for digit d, respectively. Using a Newton–Raphson method and iterating equation (A 2) until convergence, we have calculated α for each sample $[1,N]$.

(ii) *Statistical tests*

Typically, the χ^2 goodness-of-fit test has been used in association with Benford's law (Nigrini 2000). Our null hypothesis here is that the sequence of primes follow a GBL. The test statistic is

$$\chi^2 = M\sum_{d=1}^{9}\frac{(P(d)-P^e(d))^2}{P^e(d)}, \tag{A 3}$$

where M denotes the number of primes in $[1,N]$. Since we are computing parameter $\alpha(N)$ using the mean of the distribution, the test statistic follows a χ^2-distribution with $9-2=7$ degrees of freedom, so the null hypothesis is rejected if $\chi^2>\chi^2_{a,7}$, where a is the level of significance. The critical values for the 10, 5 and 1 per cent are 12.02, 14.07 and 18.47, respectively. As we can see in table 3, despite the excellent visual agreement (figure 1), the χ^2-statistic goes up with sample size and, consequently, the null hypothesis cannot be rejected only for relatively small sample sizes $N<10^9$. As a matter of fact, the χ^2-statistic suffers from the excess power problem on the basis that it is size sensitive: for huge datasets, even quite small differences are statistically significant (Nigrini 2000). A second alternative is to use the standard Z-statistics to test significant differences. However, this test is also size dependent and hence registers the same problems as χ^2 for large samples. Owing to these facts, Nigrini (2000) recommends for Benford analysis a distance measure test called mean absolute deviation (MAD). This test computes the average of the nine absolute differences between the empirical proportions of a digit and the ones expected by the GBL. That is,

Table 4. Table gathering the values of the following statistics: χ^2, maximum absolute deviation (m), MAD and correlation coefficient (r) between the observed first significant digit frequency in the M zeros in the interval $[1,N]$ and the expected generalized Benford distribution (equation (3.4) with an exponent $\alpha(N)$ given by equation (3.2) with $a=2.92$). (While χ^2-test rejects the hypothesis for very large samples due to its size sensitivity, every other test cannot reject it, enhancing the goodness of fit between the data and the generalized Benford distribution.)

N	M=no. of zeros	χ^2	m	MAD	r
10^3	649	0.14	0.32×10^{-2}	0.13×10^{-2}	0.99701
10^4	10142	0.23	0.11×10^{-2}	0.41×10^{-3}	0.99943
10^5	138069	0.75	0.54×10^{-3}	0.20×10^{-3}	0.99974
10^6	1747146	3.6	0.34×10^{-3}	0.13×10^{-3}	0.99983
10^7	21136126	20.3	0.23×10^{-3}	0.86×10^{-4}	0.99988

$$\mathrm{MAD} = \frac{1}{9}\sum_{d=1}^{9} |\, P(d) - P^e(d) \,|\,. \tag{A 4}$$

This test overcomes the excess power problem of χ^2 as long as it is not influenced by the size of the dataset. While MAD lacks a cut-off level, Nigrini (2000) suggests that the following guidelines for measuring conformity of the first digits to Benford's law: MAD between 0 and 0.4×10^{-2} implies close conformity; from 0.4×10^{-2} to 0.8×10^{-2} acceptable conformity; from 0.8×10^{-2} to 0.12×10^{-1} marginally acceptable conformity; and, finally, greater than 0.12×10^{-1}, non-conformity. Under these cut-off levels, we cannot reject the hypothesis that the first-digit frequency of the prime number sequence follows a GBL. In addition, the maximum absolute deviation m defined as the largest term of MAD is also shown in each case.

As a final approach to testing for a similarity between the two histograms, we can check the correlation between the empirical and theoretical proportions by the simple regression correlation coefficient r in a scatterplot. As we can see in table 3, the empirical data are highly correlated with a generalized Benford distribution.

The same statistical tests have been performed for the case of the non-trivial Riemann zeta zeros sequence (table 4), with similar results.

(*c*) *Cramér's model*

The prime number distribution is deterministic in the sense that primes are determined by precise arithmetic rules. However, its apparent local randomness has suggested several stochastic interpretations. Concretely, Cramér (1935, see also Tenenbaum & France 2000) defined the following model: assume that we have a sequence of urns $U(n)$, where $n=1, 2, \ldots$, and put black and white balls in each urn such that the probability of drawing a white ball in the k^{th}-urn goes as $1/\log k$. Then, in order to generate a sequence of pseudo-random prime numbers, we need only to draw a ball from each urn: if the drawing from the k^{th}-urn is white, then k will be labelled as a pseudo-random prime. The prime number sequence can indeed be understood as a concrete realization of this stochastic process. With such a model, Cramér studied, among others, the distribution of gaps between primes and the distribution of twin primes as far as statistically

Table 5. Table gathering the values of the following statistics: χ^2, maximum absolute deviation (m), MAD and correlation coefficient (r) between the observed first significant digit frequency in the Cramér model for M pseudo-random primes in the interval $[1,N]$ and the expected generalized Benford distribution (equation (3.1) with an exponent $\alpha(N)$ given by equation (3.2) with $a=1.1$).

N	M=no. of pseudo-random primes	χ^2	m	MAD	r
10^4	1189	1.20	0.17×10^{-1}	0.92×10^{-2}	0.639577
10^5	9673	0.43	0.33×10^{-2}	0.21×10^{-2}	0.969031
10^6	78693	0.39	0.59×10^{-3}	0.14×10^{-2}	0.990322
10^7	664894	0.09	0.23×10^{-3}	0.99×10^{-4}	0.999626
10^8	5762288	0.24	0.15×10^{-3}	0.53×10^{-4}	0.999855
10^9	50850064	1.23	0.11×10^{-3}	0.42×10^{-4}	0.999892
10^{10}	455062569	6.84	0.90×10^{-4}	0.33×10^{-4}	0.999914
10^{11}	4118136330	41.0	0.73×10^{-4}	0.27×10^{-4}	0.999937

speaking, these distributions should be similar to the pseudo-random ones generated by his model. Quoting Cramér: 'With respect to the ordinary prime numbers, it is well known that, roughly speaking, we may say that the chance that a given integer n should be a prime is approximately $1/\log n$. This suggests that by considering the following series of independent trials we should obtain sequences of integers presenting a certain analogy with the sequence of ordinary prime numbers P'_n.'

In this work, we have simulated a Cramér process, in order to obtain a sample of pseudo-random primes in $[1,10^{11}]$. Then, the same statistics performed for the prime number sequence have been realized in this sample. The results are summarized in table 5. We can observe that the Cramér's model reproduces the same behaviour, namely (i) the first-digit distribution of the pseudo-random prime sequence follows a GBL with a size-dependent exponent that follows equation (3.2). (ii) The number of pseudo-random primes found in each decade matches, statistically speaking, the actual number of primes. (iii) The χ^2-test evidences the same problems of power for large datasets. Bearing in mind that the sample elements in this model are independent (which is not the case in the actual prime sequence), we can confirm that the rejection of the null hypothesis by the χ^2-test for huge datasets is not related to a lack of data independence but more likely to the test's size sensitivity. (iv) The rest of the statistical analysis is similar to the one previously performed in the prime number sequence.

References

Benford, F. 1938 The law of anomalous numbers. *Proc. Am. Philos. Soc.* **78**, 551–572.

Berger, A., Bunimovich, L. A. & Hill, T. P. 2005 One-dimensional dynamical systems and Benford's law. *Trans. Am. Math. Soc.* **357**, 197–220. (doi:10.1090/S0002-9947-04-03455-5)

Berry, M. V. & Keating, J. P. 1999 The Riemann zeta-zeros and eigenvalue asymptotics. *SIAM Rev.* **41**, 236–266. (doi:10.1137/S0036144598347497)

Bogomolny, E. 2007 Riemann zeta functions and quantum chaos. *Prog. Theor. Phys. Suppl.* **166**, 19–44. (doi:10.1143/PTPS.166.19)

Chernoff, P. R. 2000 A pseudo zeta function and the distribution of primes. *Proc. Natl Acad. Sci. USA* **97**, 7697–7699. (doi:10.1073/pnas.97.14.7697)

Cramér, H. 1935 Prime numbers and probability. *Skand. Mat. Kongr.* **8**, 107–115.
Diaconis, P. 1977 The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5**, 72–81. (doi:10.1214/aop/1176995891)
Dickson, L. E. 2005 *History of the theory of numbers divisibility and primality*, vol. I. New York, NY: Dover Publications.
Edwards, H. M. 1974 *Riemann's zeta function*. New York, NY; London, UK: Academic Press.
Flehinger, B. J. 1966 On the probability that a random integer has initial digit A. *Am. Math. Mon.* **73**, 1056–1061. (doi:10.2307/2314636)
Green, B. & Tao, T. 2008 The primes contain arbitrary long arithmetic progressions. *Ann. Math.* **167**, 481–547.
Guy, R. K. 2004 *Unsolved problems in number theory*, 3rd edn. New York, NY: Springer.
Hawkins, D. 1957 The random sieve. *Math. Mag.* **31**, 1–3.
Hill, T. P. 1995*a* Base-invariance implies Benford's law. *Proc. Am. Math. Soc.* **123**, 887–895. (doi:10.2307/2160815)
Hill, T. P. 1995*b* A statistical derivation of the significant-digit law. *Stat. Sci.* **10**, 354–363.
Hill, T. P. 1996 The first-digit phenomenon. *Am. Sci.* **86**, 358–363.
Hlawka, E. 1984 *The theory of uniform distribution*. Zurich, Switzerland: AB Academic Publishers pp. 122–123.
Hürlimann, W. 2006 Benford's law from 1881 to 2006: a bibliography. (http://arxiv.org/abs/math/0607168)
Knuth, D. 1997 *The art of computer programming Seminumerical algorithms*, vol. 2. Reading, MA: Addison-Wesley.
Kriecherbauer, T., Marklof, J. & Soshnikov, A. 2001 Random matrices and quantum chaos. *Proc. Natl Acad. Sci. USA* **98**, 10 531–10 532. (doi:10.1073/pnas.191366198)
Leemis, L. M., Schmeiser, W. & Evans, D. L. 2000 Survival distributions satisfying Benford's law. *Am. Stat.* **54**, 236–241. (doi:10.2307/2685773)
Manrubia, S. C. & Zanette, D. H. 1999 Stochastic multiplicative processes with reset events. *Phys. Rev. E* **59**, 4945–4948. (doi:10.1103/PhysRevE.59.4945)
Mebane Jr, W. R. 2006 Detecting attempted election theft: vote counts, voting machines and Benford's law. In *Annual Meeting of the Midwest Political Science Association, Palmer House, Chicago, IL, 20–23 April 2006*. See http://macht.arts.cornell.edu/wrm1/mw06.pdf.
Miller, S. J. & Kontorovich, A. 2005 Benford's law, values of L-functions and the $3x+1$ problem. *Acta Arithmetica* **120**, 269–297. (doi:10.4064/aa120-3-4)
Miller, S. J. & Takloo-Bighash, R. 2006 *An invitation to modern number theory*. Princeton, NJ: Princeton University Press.
Newcomb, S. 1881 Note on the frequency of use of the different digits in natural numbers. *Am. J. Math.* **4**, 39–40. (doi:10.2307/2369148)
Nigrini, M. J. 2000 *Digital analysis using Benford's law*. Vancouver, BC: Global Audit Publications.
Nigrini, M. J. & Miller, S. J. 2007 Benford's law applied to hydrological data–results and relevance to other geophysical data. *Math. Geol.* **39**, 469–490. (doi:10.1007/s11004-007-9109-5)
Pietronero, L., Tossati, E., Tossati, V. & Vespignani, A. 2001 Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. *Physica A* **293**, 297–304. (doi:10.1016/S0378-4371(00)00633-6)
Pinkham, R. S. 1961 On the distribution of first significant digits. *Ann. Math. Stat.* **32**, 1223–1230. (doi:10.1214/aoms/1177704862)
Raimi, R. A. 1976 The first digit problem. *Am. Math. Mon.* **83**, 521–538. (doi:10.2307/2319349)
Ribenboim, P. 2004 *The little book of bigger primes*, 2nd edn. New York, NY: Springer.
Stein, M. L., Ulam, S. M. & Wells, M. B. 1964 A visual display of some properties of the distribution of primes. *Am. Math. Mon.* **71**, 516–520. (doi:10.2307/2312588)
Tenenbaum, G. & France, M. M. 2000 *The prime numbers and their distribution*. Providence, RI: American Mathematical Society.
Withney, R. E. 1972 Initial digits for the sequence of primes. *Am. Math. Mon.* **79**, 150–152. (doi:10.2307/2316536)