Physica A 440 (2015) 147–154

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# The first-digit frequencies in data of turbulent flows

Damien Biau

*DynFluid Laboratory, Arts et Métiers ParisTech, 151 Boulevard de l'Hôpital, 75013 Paris, France*

## HIGHLIGHTS

- Analysis of first significant digit distributions in data sets of turbulent flows.
- Conformance measured with Shannon's entropy.
- Results in agreement with Newcomb–Benford's law.
- The discrepancy is related with the phenomenon of intermittency in turbulence.
- A matlab program is provided in appendix.

## ARTICLE INFO

*Article history:*
Received 29 April 2015
Received in revised form 13 July 2015
Available online 21 August 2015

*Keywords:*
Turbulent flow
First significant digit
Shannon's entropy
Dissipation
Spectral method

## ABSTRACT

Considering the first significant digits (noted $d$) in data sets of dissipation for turbulent flows, the probability to find a given number ($d = 1$ or 2 or ...9) would be 1/9 for a uniform distribution. Instead the probability closely follows Newcomb–Benford's law, namely $P(d) = \log(1 + 1/d)$. The discrepancies between Newcomb–Benford's law and first-digits frequencies in turbulent data are analysed through Shannon's entropy. The data sets are obtained with direct numerical simulations for two types of fluid flow: an isotropic case initialized with a Taylor–Green vortex and a channel flow. Results are in agreement with Newcomb–Benford's law in nearly homogeneous cases and the discrepancies are related to intermittent events. Thus the scale invariance for the first significant digits, which supports Newcomb–Benford's law, seems to be related to an equilibrium turbulent state, namely with a significant inertial range. A matlab/octave program provided in appendix is such that part of the presented results can easily be replicated.



## 1. Introduction

In 1881 Newcomb [1] observed a rather strange fact: tables of logarithms in libraries tend to be quite dirty at the beginning and progressively cleaner throughout. This seemed to indicate that people had more occasions to calculate with numbers beginning with 1 than with other digits. Newcomb concluded that the frequency of the leftmost, nonzero digit $d$ closely follows the probability law:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right) \quad d = 1, 2, \ldots 9. \tag{1}$$

Hence the numbers in typical statistics should have a first digit of 1 about 30% of the time, but a first digit of 9 only about 4.6% of the time (the other values can be found in Table 1). Also, the probability that the first significant digit is an odd number is 60%. This formula is also valid for the digits beyond the first, for example the distribution of the $k$ first digits is given by Eq. (1) with $d = 10^{k-1}, \ldots 10^k - 1$.

*E-mail address:* damien.biau@ensam.eu.

http://dx.doi.org/10.1016/j.physa.2015.08.016
0378-4371/

Newcomb's article went unnoticed until 1938 when Benford [2] has independently followed the same path: starting from observations about logarithm books, he deduced the same law. Benford also noted that this law is base-invariant, the definition (1) is given here with a decimal representation (base-10) for convenience. Until today this empirical law has been tested against various data sets ranging from mathematical curiosity to natural sciences [3,4].

Then there have been many attempts to rationalize this empirical law [5–10]. Recently Fewster [10] has provided an intuitive explanation based on the fact that any distribution has tendency to satisfy Newcomb–Benford's law, as long as the distribution spans several orders of magnitude and as long as the distribution is reasonably smooth.

Pietronero et al. [7] have shown that Newcomb–Benford's law is equivalent to the scale invariance of data set. Multiplying a data set $X$ by a factor $\lambda$ rescales the first-digits distribution such that: $P(d[\lambda X]) = \lambda^m P(d[X])$, where $d[X]$ denotes the first significant digit of the variable $X$ and $m$ is an exponent to be determined. The general solution of this equation takes the form of power-law. The probability distribution $P(d)$, namely the sub-interval of $[1, 10)$ occupied by the first digit $d$, is obtained by integration:

$$P(d) = K \int_d^{d+1} z^{-\alpha}\, \mathrm{d}z$$

where $K$ is a necessary constant to enforce the condition $\sum P = 1$. Newcomb–Benford's law is recovered for $\alpha = 1$, while $\alpha \neq 1$ leads to the generalized distribution found by Pietronero et al. [7]:

$$P(d) = \frac{(d+1)^{1-\alpha} - d^{1-\alpha}}{10^{1-\alpha} - 1}. \tag{2}$$

The generalized Newcomb–Benford law has been applied to the distribution of leading digits in the prime number sequence by Luque and Lacasa [11], the results have shown an asymptotic evolution towards the uniform distribution ($\alpha = 0$). Note that as particular case $\alpha = 1$, Newcomb–Benford's law satisfies a strong invariance, *i.e.* $P(d[\lambda X]) = P(d[X])$. This result was first demonstrated by Hill [6]. As illustration, we consider $X$ such that $d[X] = 7$ and $\lambda = 2$ as scaling factor, thus $d[\lambda X] = 1$, but the secondary digit can only be 4 or 5. With the relation (1), it is easy to check the scale-invariance for this example: $P(7) = P(14) + P(15)$.

Self-similarity is also a well explored topic in turbulent flows. The first step was made by Richardson in the late 19th century with the assumption of a universal cascade process of the energy: the energy input at large scale is successively transferred to finer eddies. This idea has been refined by Kolmogorov and Obukhov, the cascade is then assumed to occur in a space-filling, self-similar way. Formally there should be a unique scaling exponent for the structure functions $S(r)$ such that $S(\lambda r) = \lambda^h S(r)$. Today the departures from the Kolmogorov scaling prediction are identified to different causes: the Reynolds number is not large enough, the flow is not exactly isotropic or the self-similarity assumption is not valid. The last point is related to the intermittent feature of the cascade process, resulting in anomalous scaling, or no unique scaling exponent. Further information can be found in the book by Frisch [12].

In the present work, the distribution of first significant digits is used as an alternative statistical tool for analysing turbulent flows. Newcomb–Benford's law is compared to data sets coming from numerical simulations of the Taylor–Green vortices (homogeneous case) and the plane Poiseuille flow (inhomogeneous case). In the next section the numerical method is presented and the Shannon entropy is introduced as a diagnostic tool. Results are presented in the following section. The validity and the possible extensions of the method are discussed in the last section.

## 2. Method

The momentum and mass conservation equations applied to the fluid lead to the Navier–Stokes equations, written in non dimensional form:

$$\begin{aligned} &\nabla \cdot \mathbf{u} = 0, \\ &\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + Re^{-1} \nabla^2 \mathbf{u} \end{aligned} \tag{3}$$

the flow is assumed incompressible and the Reynolds number ($Re$) is the only control parameter.

It should be noted that the quadratic nonlinear term in the Navier–Stokes equations can be related to Newcomb–Benford's law. If we assume a scale invariant data set, then the peculiar Newcomb–Benford's law could be rooted in this non-linear term. Let consider the quadratic transformation, $X \to X^2$, mimicking the nonlinear convective term in Navier–Stokes equations (6). The probability to find the first digit of $X$ in the interval $[d, d+1)$ equals the probability to find the first digit of $X^2$ in $[d^2, (d+1)^2)$:

$$\int_{d^2}^{(d+1)^2} z^{-\tilde{\alpha}}\, \mathrm{d}z = \int_d^{d+1} z^{-\alpha}\, \mathrm{d}z$$

leading to $\tilde{\alpha} = (\alpha + 1)/2$. For any distribution, characterized by an exponent $\alpha$, the quadratic transformation generates a distribution with an exponent $\tilde{\alpha}$ which is closer to 1 than initial $\alpha$. As a consequence the quadratic non-linear term