Zeitschrift: Elemente der Mathematik
Herausgeber: Schweizerische Mathematische Gesellschaft
Band: 72 (2017)
Heft: 2

Buchbesprechung: Rezensionen

Elem. Math. 72 (2017) 87 – 88
0013-6018/17/020087-2
DOI 10.4171/EM/330

Elemente der Mathematik

Rezensionen

A. Berger and T.P. Hill: An Introduction to Benford's Law. viii+248 pages, \$75.00. Princeton University Press, Princeton and Oxford, 2015; ISBN 978-0-691-16306-2 (hard cover).

S.J. Miller (ed.): Benford's Law. Theory and Applications. xxvi+438 pages, \$75.00. Princeton University Press, Princeton and Oxford, 2015; ISBN 978-0-691-14761-1 (hard cover).

The present review is double for two reasons: firstly both books deal with the same subject and mostly they are complementary in some sense. The so-called Benford Law (hereafter BL) is the observation that, in many diverse collections of numerical data, the occurrence of the first digit is not uniform. Indeed, the digit 1 appears more often than 2 as leading digit, in turn, 2 appears more frequently than 3 and so on, the digit 9 appearing approximately only 5% of the time. More precisely, a set of numerical data satisfies BL if, for every digit $d \in \{1, 2, \dots, 9\}$, the probability that the first digit be d is equal to $\log(1 + 1/d)$ (here log denotes the base-10 logarithm). Thus the probability that the first digit be 1 equals approximately 0.301, the probability that the first digit be 2 is approximately equal to 0.176 and so on.

The law was discovered first in 1881 by the American astronomer Simon Newcomb but his article passed completely unnoticed. Newcomb observed that the pages of logarithms tables were more worn in the first digits than in the last ones, and after some heuristics, he deduced that occurrence of the first digits should follow the logarithmic law stated above. BL was re-discovered, apparently independently, by Frank Benford in 1938. He gathered 20,229 entries from 20 diverse data sources and he noticed that, even if some data sources did not follow BL too well, the whole set of data did in a rather satisfactory way.

Let us comment on and present a selection of the topics in Berger's and Hill's book first. Printed on glazed paper and containing many colored figures, it is an up-to-date and detailed introduction to BL, containing almost all detailed proofs and many illuminating examples, but with only a tiny part (eight pages) devoted to a rough description of applications of BL. After a brief historical sketch, the authors introduce the main tool that is necessary to define BL in a mathematically satisfactory way, namely the *significand σ-algebra $\mathcal{S}$* which is the σ-algebra on $\mathbb{R}^+$ generated by the *significand function* $S : \mathbb{R} \to [1, 10)$ defined as follows: $S(0) = 0$, and, if $x \neq 0$, $S(x)$ is the unique number $t \in [1, 10)$ such that $|x| = t \cdot 10^k$ for some (unique) $k \in \mathbb{Z}$. (Notice that only the usual base 10 is considered.) Thus, $\mathcal{S}$ is the set of all Borel subsets A of $\mathbb{R}^+$ of the form

$$\bigcup_{k \in \mathbb{Z}} 10^k B$$

where B is any Borel set of $[1, 10)$. Consequently, the *Benford distribution* $\mathbb{B}$ is the unique probability measure on $(\mathbb{R}^+, \mathcal{S})$ such that

$$\mathbb{B}(S \leq t) = \mathbb{B}\left(\bigcup_{k \in \mathbb{Z}} 10^k [1, t]\right) = \log(t), \quad t \in [1, 10).$$

In order to fix ideas, we focus now on sequences. A sequence $(x_n)_{n \geq 1} \subset \mathbb{R}$ is *Benford* if

$$\lim_{N \to \infty} \frac{\#\{1 \leq n \leq N : S(x_n) \leq t\}}{N} = \log(t) \quad (t \in [1, 10)).$$

Observe that it implies F. Benford's first definition above since $|x|$ has first digit d if and only if $d \leq S(x) < d+1$. In the same vein, *Benford functions* and *Benford random variables* are defined analogously.

It is worth mentioning that BL for sequences can be rephrased as follows: *A sequence* (x_n) *is Benford if and only if* $(\log(|x_n|))$ *is uniformly distributed modulo* 1 *in H. Weyl's sense.* This observation allows using many results on uniformly distributed sequences. For instance, a geometric sequence (r^n) is Benford if and only if $\log(|r|)$ is irrational. Similarly, BL is *scale-invariant* in the following sense: a sequence (x_n) is Benford if and only if (ax_n) is Benford for every number $a > 0$.

The authors also study many real-valued deterministic processes, *i.e.*, sequences (x_n) such that $x_{n+1} = f(x_n)$ for a suitable function f. For instance, let $g : I \to \mathbb{R}$ be a differentiable function defined on some interval I. Then Newton's method is such a process associated to the function $f(x) = x - \frac{g(x)}{g'(x)}$. Under suitable conditions, it is proved that if $x^* \in I$ is a zero of g, then the sequences $(x_n - x^*)$ and $(x_{n+1} - x_n)$ are Benford for almost all initial values near x^*.

Relatively important examples of Benford sequences are those that satisfy some linear difference equations (e.g., the Fibonacci sequence). The authors base their study on *multi-dimensional linear processes*, *i.e.*, processes given by primitive square matrices A whose spectral radius ρ is such that $\log(\rho) \notin \mathbb{Q}$.

Berger's and Hill's monograph is very clearly written and it should be a reference book on Benford's Law. However, I regret the absence of several interesting results from articles written in other languages than english (in particular very interesting articles in french from J.-P. Delahaye and N. Gauvrit) as well as their references.

Let us turn to Miller's book now. On the contrary of Berger's and Hill's, chapters are written by different authors, Part I being more or less a summary of the latter, and a half of the monograph being devoted to applications of BL.

Unfortunately, the quality of chapters in Miller's book varies a lot: if Parts I and II are very well written, it is not the case in several chapters of Parts III to VI, and it is unfortunate, as these parts are devoted to applications. For instance, Chapter 8, by M. Nigrini, is very clearly written and fascinating, Chapter 19, by Miller *et al.*, is of rather high quality, but many other chapters seem to be directly taken from proceedings of the 2007 *Conference on the Theory and Applications of Benford's Law*, which was organized by some of the authors in Santa Fe, NM. Indeed, notation are not unified, there are many useless repetitions of basic definitions and some prerequisites in various fields are missing.

Let us describe briefly its content. Part I contains an abstract of Berger's and Hill's book in the first two chapters, and Chapter 3 discusses some applications of Fourier analysis to BL. Part II discusses distributions that follow *approximately* BL; for instance, Chapter 5 by Dümbgen and Leuenberger derive helpful bounds from the total variation of the density function. That part ends with a long and technical chapter on Lévy processes. Next, as mentioned above, Parts III to VI are devoted to diverse applications of BL. Part III presents some topics on the rather convincing application of BL to accounting on the one hand, and much less convincing applications to vote fraud on the other. Part IV is devoted to applications to Economics: an example is given by European statistics where the authors of Chapter 11 claim that some countries of the euro zone such as Greece attempt to manipulate their fiscal monitoring data in order to meet the criteria of the Stability and Growth Act, and compare them to social data sets of the same country; the result is that the former do not satisfy BL while the latter do. Parts V and VI present applications of BL to natural sciences and to image processing respectively. In Chapter 16, D. Hoyle describes in a clear way the occurence of BL in natural sciences ranging from half-lives of elementary particles to statistical physics and to biosciences. Concerning imaging systems, Chapter 18 on medical imaging is rather obscure, contrary to the next one which describes applications of a slightly more general form of BL to steganography. Finally, the last part is a chapter devoted to exercises related to the topics of the former chapters. Their levels of interest and difficulty vary from one another.

In conclusion, reading both books would give to the reader an excellent overview of what BL is, but the reader interested in what BL is good for should spend much effort in the search of information referred to in the related chapters of Miller's book.

Paul Jolissaint, Porrentruy

Typesetting & Layout: *mathScreen online, CH-4123 Allschwil*