Forum on Benford's law and statistical methods for the detection of frauds

Lucio Barabesi[1] · Andrea Cerioli[2] · Domenico Perrotta[3]

1 Introduction to the forum

This forum intends stimulating cross-domain research in Benford's law theory and robust statistics. Our point of view is introduced in Sect. 2, with the intention to show the rationale and benefit of extending to the statistical community the discussions on the probabilistic and number-theoretic formulations of Benford's theory. The point of contact between the two scientific disciplines is anti-fraud. We use international trade as a specific motivating domain, for its great societal, economic and policy impact, but especially because it proved to benefit concretely from the cross-domain approach discussed by the forum, as Sect. 3 explains. Section 4 hitches the concrete examples of this dialogue selected by the forum.

2 The fascination of Benford's law

Benford's law is a fascinating phenomenon related to the frequency of leading digits in many real-life collections of numerical data. Benford's law is also an appropriate example of Stigler's law of eponymy (Stigler 1980), which asserts that no scientific result is named after its original discoverer. Indeed, the first known result about this law can be attributed to Simon Newcomb, the well-known mathematician and

✉ Domenico Perrotta
domenico.perrotta@ec.europa.eu

Lucio Barabesi
lucio.barabesi@unisi.it

Andrea Cerioli
andrea.cerioli@unipr.it

1   Department of Economics and Statistics, University of Siena, Siena, Italy

2   Department of Economics and Management and University Centre "Robust Statistics Academy" (Ro.S.A.), University of Parma, Parma, Italy

3   European Commission, Joint Research Centre (JRC), Ispra, Italy

astronomer, who derived it on the basis of the pattern of the first significant digits contained in logarithm tables (Newcomb 1881). Frank Benford, a physicist at the General Electric Company, rediscovered and publicized the same phenomenon without being aware of the Newcomb's finding (Benford 1938), further emphasizing that it applies to many types of numerical data, ranging from death rates to stock prices, from baseball statistics to the area of lakes. In its basic formulation Benford's law asserts that the leading digit d in a list of numbers from many real-life sources of data is not uniformly distributed, as one might naively expect. Instead, it occurs with a probability given by $\log_{10}(1 + \frac{1}{d})$ for $d = 1, \ldots, 9$. This distributional prescription, often named the "first-digit law", is a peculiar manifestation of Benford's law which, nevertheless, should not be identified with the law itself. The complete form of Benford's law has been understood many years after the publication of Benford's work and involves a more general framework based on the joint distribution of the whole set of significant digits of each number (see Berger and Hill 2011a, b, for details). From a mathematical perspective, it is astonishing that appropriate versions of Benford's law appear in number theory, such as in the Weyl's Equidistribution Theorem (Weyl 1916), or in integer sequences, such as in the celebrated Fibonacci sequence or in the factorial sequence (Diaconis 1977). The counterintuitive nature of Benford's law is emphasized by several authors in books on mathematical paradoxes; see, e.g., Chang (2012, Section 1.2.1), Gorroochurn (2012, "Benford and the peculiar behavior of the first significant digit (1938)", Chatper 27) and Havil (2008, "Benford's law", Chapter 16). Benford's law is also described as one of the 250 mathematical milestones by Pickover (2009, milestone for the year 1881, when introduced by Newcomb), while Knuth (1997, p. 254) devotes a large discussion to the issue in a book of his famous series "The Art of Computer Programming". More recently, Bijma et al. (2017) include an entire appendix (to Section 2) to Benford's law, while Demidenko (2020) provides a long discussion of Benford's law in a specific section (Section 2.12). Other amusing descriptions of Benford's law are provided by Wagon (2010, Section 21.6), Olofsson (2015, "Number one is number one", Chapter 9), Dworsky (2019, Chapter 14) and Tijms (2019, "Benford Goes to the Casino", Chapter 5).

Benford's law has attracted the attention of mathematicians, as well as of applied scientists. As a matter of fact, the large number of items collected in the exhaustive and up-to-date dedicated repository Benford Online Bibliography (Berger et al. 2009) attests the lively interest on the topic. The Benford Online Bibliography (BOB thereafter) currently contains over than 1500 entries (last accessed on 31 December 2020), encompassing books, papers and manuscripts. By browsing the archive chronologically, it is apparent that Benford's law has solely received a marginal curiosity before 1970 (less than 40 entries in BOB for this period). The topic has drawn an increasing interest in the seventies and the eighties of the past century, especially from mathematicians working in the field of probability (BOB provides about 110 items for these 2 decades). The real scientific rise of Benford's law has occurred in the nineties (about 110 entries in BOB). Surely, in that period the research on Benford's law was boosted by the articles of Ted Hill (see especially Hill 1995a, b), addressing the mathematical foundations of the law, and of Mark

Nigrini (see Nigrini 1992, and the citations to his PhD thesis), concerning its practical applications to the detection of frauds and tax evasion. In the new century, a flourish of papers devoted to Benford's law has occurred (about 1300 entries are present in BOB for the period 2000–2020). Accordingly, the various topics dealing with Benford's law have been subjected to a more systematic treatment, while at least four authoritative monographs have appeared in recent years (Berger and Hill 2015; Kossovsky 2015; Miller 2015; Nigrini 2012). The increasing trend in the bibliography of Benford's law is well documented in Fig. 1, where the time series of the item numbers per year in BOB is plotted for the period 1980–2020.

The relevance of Benford's law in scientific research is also confirmed by the number of items on this topic in Scopus, which is obviously a more selective repository. A query in Scopus on the string "Benford's law" (with quotation marks) contained in the article title, or/and in the abstract, or/and in the keywords, produces an output of 560 items (last accessed on 31 December 2020). The marked positive trend in the contributions to Benford's law is evidenced in Fig. 2, where the time series for the item numbers per year in Scopus is depicted for the period 1980–2020. Three main contributors to the present Forum, namely Arno Berger, Ted Hill and Steven Miller, are among the most prolific and cited authors in this selected database from Scopus. Noteworthy, Hill (1995b) is the most cited item in such a collection of papers. Moreover, it is amusing to see that the first-digit distribution of the number of citations to the items of the database follows Benford's law! This empirical distribution is plotted in Fig. 3, along with the Benford's first-digit
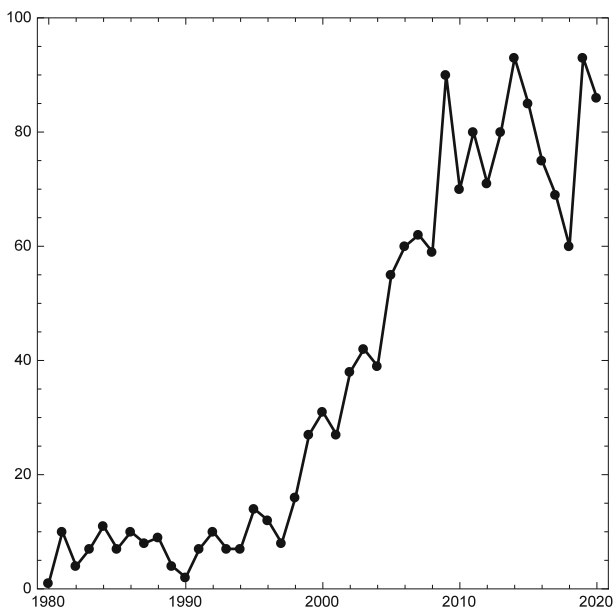
Fig. 1 Time series of the item numbers per year in the dedicated repository Benford Online Bibliography (1980–2020)
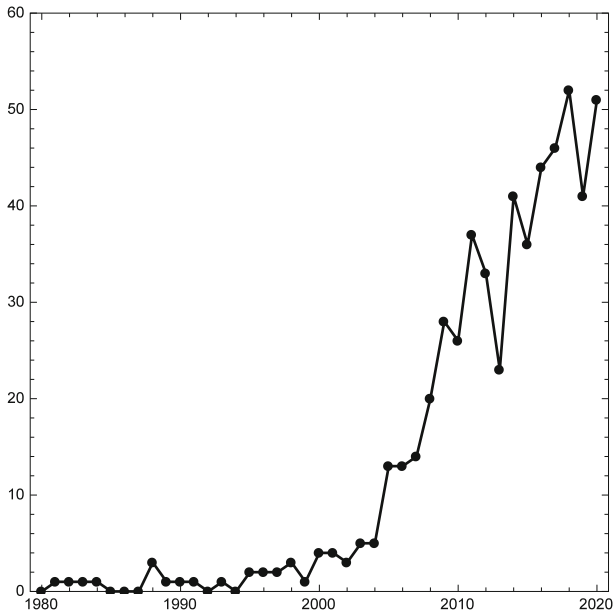
Fig. 2 Time series of the numbers (per year) of papers in Scopus with the quoted string "Benford's law" in the article title, or/and the abstract, or/and the keywords
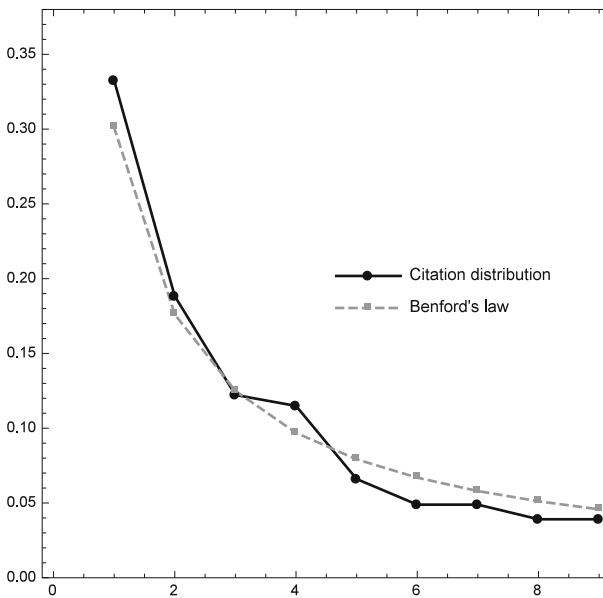
Fig. 3 First-digit distribution of the citations to papers in Scopus with the quoted string "Benford's law" in the article title, or/and in the abstract, or/and in the keywords, together with the first-digit distribution under Benford's law

distribution. The fitting is noticeable, since the value of the chi-square statistic equals 8.10 on 8 degrees of freedom.

The readers of Statistical Methods and Applications may appreciate the fact that Italian scholars lively contributed to the early developments of Benford's law, especially from the mathematical side. Indeed, some Italian scholars produced a few noticeable papers which unfortunately are barely known, since they were often written in Italian or French and were published in mathematical Italian journals (in any case, they are present in BOB). As an example, the manuscript by Herzel (1956) was one of the first papers devoted to Benford's law in the literature. This work attempts to explain the law by adopting a system of urns (with balls numbered with the first n integers) which are randomly chosen by some different schemes—a seminal idea which may resemble the formal setting proposed by Hill (1995a). Even if none of these schemes produces asymptotic results as the number of urns increases, Herzel (1956) obtained some integral approximations which are close to Benford's law.

Many of the first articles produced by the Italian school on Benford's law were developed under the interesting De Finetti's view of finitely-additive probability and by assuming the concept of a "natural density" on the integers. In this setting, pioneering papers were written by Scozzafava (1981) and Regazzini (1982). In particular, Scozzafava (1981) provided a justification of Benford's law based on the well-known concept of non-conglomerability initially proposed by Bruno De Finetti. In a similar framework, Fuchs and Letta (1984, 1996) introduced the remarkable notion of conditional density with respect to a subset of the integers. In this context, Benford's law implies that the conditional logarithmic density is equal to the (non-conditional) logarithmic density. Moreover, Fuchs and Letta (1996) showed that, for a large class of subsets of the integers, the upper and lower arithmetic and logarithmic densities coincide with the corresponding conditional densities with respect to the set of prime numbers. Such a result emphasizes that the set of prime numbers satisfies an "extended" Benford's law (see Giuliano Antonini and Grekos 2005, for further details and generalizations). Further interesting results based on the finitely-additive probability approach can be found in Candeloro (1998), together with the definition of the concept of a "Benford-compatible" random variable. We also note that Berger and Hill (2015) devote an entire chapter of their monograph to finitely-additive probability and Benford's law. We conclude this brief survey by remarking that, in the countably-additive approach, Volčič (1996) provided independently of Hill (1995a) an alternative and elegant explanation of Benford's law. Volčič (1996) reduced the sample space to the interval [1, 10) and introduced a simple scale-invariance condition from which Benford's law may be deduced. The approaches proposed by Hill (1995a) and Volčič (1996) are closely related (see the Addendum in Volčič 1996). The numerous entries related to Italian authors in BOB testify the interest received by Benford's law in the Italian research community during the last decades.

3 The challenges of fraud detection in international trade

From the point of view of applications, our main interest is the detection of the frauds that may arise in international trade to and from the European Union (EU). The fact that numbers tend to begin with lower rather than with higher digits also in companies' tax returns and financial reports, has intrigued the international Customs community, for the potential extension of Benford's law to the detection of import/export flows whose digits diverge substantially from the expected distribution. Indeed, there are a variety of frauds and irregularities associated to the miss-declaration of the value or quantity (and therefore ultimately the price) of a trade transaction, ranging from undervaluation, money laundering, VAT evasion, e-commerce, but also attempts to hide the true origin or nature of the good or to bypass trade restrictions. In all cases, the purpose is to evade import duties and related taxes on all types of commodities.

So far authors working in this specific anti-fraud domain have adapted and extended existing statistical approaches, with varying degrees of complexity and following the general target to identify suspect transactions that deviate from a reference model defined on "regular" trade, hopefully the majority of the observations. For example, Perrotta and Torti (2010), Cerioli and Perrotta (2014), Cerasa and Cerioli (2017) and Riani et al. (2018) have used robust methods to identify regression outliers and linear structures in data relating the values of the imported goods to the corresponding quantities, in order to estimate the market trade prices and the major, possibly systematic, deviations from them. Similarly, Rousseeuw et al. (2019) have worked on the identification of sudden and structural changes in time series of trade that may pinpoint suspicious transactions. Some of these authors have contributed to the present Forum, but of course there are other conventional data analysis techniques for fraud detection that could have been considered, in particular those using tools from statistical learning and knowledge discovery in data streams. Authoritative representatives of this rich literature are Bolton and Hand (2002), who applied their methods in a variety of other (non-customs) contexts, ranging from insurance claims to credit card transactions and tax return claims.

The majority of the statistical approaches currently available for the detection of frauds in international trade measure a distance of the detected anomalies from the regular part of the data, which is a proxy to the amount that has been fraudulently distorted. Therefore, they can be used on a large scale to estimate the overall potential loss for the national or EU budgets, since most of the applicable duty and tax rates are known percentages of the declared values (FISCALIS 2016). This is often done using representative data samples (bottom-up/direct methodologies) and in association with general models on the tax revenues (top-down/indirect methodologies), like in the pioneering proposal of Scala (1966) for the Italian panorama. Specifically, Scala (1966) aimed at estimating the total of tax evasion by modeling income distribution with a log-Normal distribution and by subsequently comparing the fitted distribution with the curve of collected tax distribution. The amount of tax evasion was finally estimated by means of a truncated likelihood. It is

remarkable to note the modernity and the potential relevance of such a contribution, written well before the upsurge in computing power that anti-fraud researchers experienced at the beginning of the twenty-first century. Therefore, estimation of tax gaps is manifestly an old but always relevant policy issue, which is receiving considerable political importance within the EU. It suffices to mention a recent EU Parliament resolution (European Parliament 2018), supported by a thoughtful study (European Parliament 2019), which "called the Commission to develop a suitable methodology and produce periodic estimates of the customs gap" for which, contrary to other types of tax revenues, there is no literature yet.

The advantages offered by these conventional approaches based on the relative position of the observed data become a weakness in the presence of subtle manipulations conceived to mask the values that are subject to tax impositions. For example, a common fraudulent practice on the customs values declared at import is to deftly play with the price of expensive commodities so that the corresponding transactions become hidden among the cheaper low-quality ones. Clearly, these manipulations remain undetected by robust statistical methods, as the data of both high and low quality commodities would overlap after the price manipulation. In addition, the detection of outlying flows by means of conventional approaches is often not sufficient to address complex cases involving several thousand transactions and dozens of companies in different countries, which is the typical context of international investigations operated by the European Anti-fraud Office (OLAF) of the European Commission (EC) (European Court of Auditors 2017; OLAF 2018). In fact, the identification of the criminal networks behind these cases relies much more on evidence on the typical behavior of the single operators, rather than on the precise number and severity of their anomalous transactions. Information at a trader level relies on expensive and rather rare subject matter knowledge, as it is typically gathered from Customs, tax and port authorities and from reliable economic operators across the globe. Benford's law applied to the transactions of each trader can thus offer cheap tools to reduce the scope of the search to operators that attempt to "cook the books".

Of course, there are some pitfalls to be avoided also with tools derived from Benford's law. One major issue is to adapt the basic rules in order to take into account the peculiarities of specific data domains, as what works for financial reports may not work for the customs prices observed under certain trade conditions. Then, it is crucial to make sure that the possibilities of false alarms are minimized, in order to make optimal use of the limited human resources available at Customs and related public anti-fraud services. These are some of the problems we have personally helped to solve (Barabesi et al. 2018; Cerioli et al. 2019; Barabesi et al. 2021), as part of the support to fraud detection that the EC Joint Research Centre (JRC) provides to OLAF. The JRC has also worked together with the Customs authorities of some Member States, using anonymised data to provide a preliminary test of the reliability of the suggested Benford's tools on customs declarations. By feeding these data into the Benford's methodology, not only could they confirm the findings of the authorities, but they also found more manipulated declarations than it was originally discovered. This suggests that the method can be helpful in providing

Customs authorities with evidence of potential fraud among traders not previously classified as fraudsters or even not considered as suspicious.

The EC is now well placed to experiment the anti-fraud approach based on Benford's law on a much larger scale, as it disposes of a "Customs Surveillance system" that centralises all EU import and export declarations collected from the national Customs authorities on a daily basis (Perrotta et al. 2020). Other related financial databases, complemented by appropriate detection tools, are or will be soon available to the EC, including one aimed at pinpointing suspicious VAT activities and another focusing on relevant cross-border payments. The remarkable increase in data availability is opening new extraordinary opportunities to fight against financial frauds in the EU. We believe that the present Forum could become a flash point in the modernization process of the EU anti-fraud services, of which the solid collaboration between OLAF and JRC has been integral part for more than 20 years. Therefore, we are extremely grateful to all the contributors to the Forum because they have made a tangible example of how much the academic statistical community is increasingly and actively involved in such a modernization process.

4 Contributions to the forum

The Forum hosts five papers, which nicely fits our limited but meaningful aim to stimulate, in a future-looking perspective, a reflection on the cross-domain potential of the statistical tools derived from Benford's law and its extensions, such as the generalized Benford's law (Barabesi and Pratelli 2020). Through dissemination of the mathematical foundations of this research path, we hope that the papers will contribute to foster the applicability of principled statistical methodologies to major anti-fraud problems arising both inside and outside the customs and financial domains. We can also see a concrete indication emerging from the five papers, pointing to the great potential of integrating fraud signals derived from the use of Benford's law with those obtained under alternative, and more established, approaches based on robust statistics and business analytics.

The leading paper, by world-class experts Berger and Hill (2020), is directed to a wide audience of statisticians and gives a concise—yet accurate—survey of the main theoretical issues related to Benford's law. Their review also includes many explanatory examples and the statements of the main theorems. Moreover, the authors report a comprehensive collection of useful references and show how to avoid some common pitfalls due to a naive understanding of the topic. We thus believe that this work has the right credentials to become a very helpful starting point for all the scholars who are eager to delve into the theory and the applications of Benford's law.

The second article, by Farris et al. (2020) is rather technical and deals with the connection between recurrence relations and Benford's law. More precisely, after emphasizing that recurrences with constant coefficients accomplish Benford's law, the authors extend the results already known in the literature to linear recurrence relations with non-constant coefficients, as well as to higher-degree recurrences and

multiplicative recurrences with non-constant coefficients. These findings are even considered in the environment of stochastic recurrence relations.

The third paper, by Mumic and Filzmoser (2021), proposes an interesting multivariate approach to test whether the observed first-digit frequencies follow the theoretical Benford's distribution. The approach is novel and relies on the concept of compositional data, which examines the relative information among the frequencies with which the different values of the leading digit occur. An application to the problem of auditing for music streaming data is also considered, thus providing a nice connection between the two main topics of this Forum.

The last two papers address important issues in the practice of fraud detection. One of these is that in labeled data the fraudulent and genuine classes are typically very unbalanced, so that classifiers tend to favour the genuine group. The article by Baesens et al. (2021) proposes to tackle this problem with a robust version of an oversampling bootstrap technique that creates synthetic data mimicking the fraudulent class, while also taking the potential presence of outliers into account, especially if they occur in the minority class. Their approach allows both to understand why an observation is flagged as suspicious and to compute accurate performance measures on the imbalanced classes. Both features appear to be important in contexts where statistical evidence is used in Court, which should be the spotlight in anti-fraud research.

The ultimate intention of the final paper by Torti et al. (2021) is similar, although framed in a different methodological context: robust regression clustering. In fact, the contribution addresses interconnected problems that are scientifically interesting and also practicable in applications. The authors focus on the optimal choice of hyper-parameters and tuning constants in robust clustering model of actual use, such as the number of groups, the level of trimming and the scatter constraints, and apply their methods to international trade data. Again, the practical point here is that unstable or heuristic choices of such crucial model features would not be justifiable in Court. Nor would it be justifiable to change approach and model choices depending on the dataset under scrutiny: a coherent and sufficiently general approach is what this contribution tries to provide. With this effort, the authors bring ahead ideas that were discussed in recent years by an issue of this journal addressing the general theory of monitoring (Cerioli et al. 2018) covering, among different methodological and practical issues, also a customs-fraud application (Perrotta and Torti 2018).

We close wishing that this Forum will contribute to trigger the setting of gold standards for statistical anti-fraud analysis, with Benford's analysis as one of the proposed components.

episode of *Connected: The Hidden Science of Everything*). Last, but not least, we are greatly indebted to the former and to the present Editor of Statistical Methods and Applications, Professors Tommaso Proietti and Carla Rampichini, for accepting with enthusiasm our request to share the benefits of Benford's law and the challenges of fraud detection with the readers of this Journal. The authors line of work on Benford's law has been supported by: (1) the JRC's first Work Programme for 2014–2015 under Horizon 2020, through the institutional research line of the JRC's Text and Data Mining Unit and a Proof of Concept supported by the JRC's Technology Transfer Office; (2) the Hercule 3 Anti-fraud Programme of the European Union, managed by OLAF; (3) the Programme "FIL-Quota Incentivante" of University of Parma and co-sponsored by Fondazione Cariparma.

References

Baesens B, Höppner S, Ortner I, Verdonck T (2021) robROSE: a robust approach for dealing with imbalanced data in fraud detection. Stat Methods Appl. https://doi.org/10.1007/s10260-021-00573-7

Barabesi L, Pratelli L (2020) On the generalized Benford law. Stat Probab Lett 160:1–283

Barabesi L, Cerasa A, Cerioli A, Perrotta D (2018) Goodness-of-fit testing for the Newcomb–Benford law with application to the detection of customs fraud. J Bus Econ Stat 36:346–358

Barabesi L, Cerasa A, Cerioli A, Perrotta D (2021) On characterizations and tests of Benford's law. J Am Stat Assoc. https://doi.org/10.1080/01621459.2021.1891927

Benford F (1938) The law of anomalous numbers. Proc Am Philos Soc 78:551–572

Berger A, Hill TP (2011a) A basic theory of Benford's law. Prob Surv 8:1–126

Berger A, Hill TP (2011b) Benford's law strikes back: no simple explanation in sight for mathematical gem. Math Intell 33:85–91

Berger A, Hill TP (2015) An introduction to Benford's law. Princeton Univ. Press, Princeton

Berger A, Hill T (2020) The mathematics of Benford's law: a primer. Stat Methods Appl. https://doi.org/10.1007/s10260-020-00532-8

Berger A, Hill TP, Rogers E (2009) Benford online bibliography. http://www.benfordonline.net. Accessed 31 Dec 2020

Bijma F, Jonker M, van der Vaart A (eds) (2017) An introduction to mathematical statistics. Amsterdam University Press, Amsterdam

Bolton RJ, Hand DJ (2002) Statistical fraud detection: a review. Stat Sci 17:235–255

Candeloro D (1998) Some remarks on the first digit problem. Atti del Seminario Matematico e Fisico dell'Università di Modena XLVI:511–532

Cerasa A, Cerioli A (2017) Outlier-free merging of homogeneous groups of pre-classified observations under contamination. J Stat Comput Simul 15:2997–3020

Cerioli A, Perrotta D (2014) Robust clustering around regression lines with high density regions. Adv Data Anal Classif 8:5–26. https://doi.org/10.1007/s11634-013-0151-5

Cerioli A, Riani M, Atkinson A, Corbellini A (2018) The power of monitoring: how to make the most of a contaminated multivariate sample. Stat Methods Appl 27:559–587

Cerioli A, Barabesi L, Cerasa A, Menegatti M, Perrotta D (2019) Newcomb–Benford law and the detection of frauds in international trade. Proc Natl Acad Sci USA 116:106–115

Chang M (2012) Paradoxes in scientific inference. CRC Press, Boca Raton

Demidenko E (2020) Advanced statistics with applications in R. Wiley, New York

Diaconis P (1977) The distribution of leading digits and uniform distribution mod 1. Ann Probab 5:72–81

Dworsky L (2019) Probably not, 2nd edn. Wiley, Hoboken

European Court of Auditors (2017) Import procedures: shortcomings in the legal framework and an ineffective implementation impact the financial interests of the EU. https://www.eca.europa.eu/Lists/ECADocuments/SR17_19/SR_CUSTOMS_EN.pdf, special Report No 19/2017 (pursuant to Article 287(4), second subparagraph, TFEU)

European Parliament (2018) Fighting customs fraud and protecting EU own resources (2018/2747(RSP)). https://www.europarl.europa.eu/doceo/document/TA-8-2018-0384_EN.html

European Parliament (2019) Protection of EU financial interest on customs and VAT: cooperation of national tax and customs authorities to prevent fraud. https://doi.org/10.2861/428486

Farris M, Luntzlara N, Miller SJ, Shao L, Wang M (2020) Recurrence relations and Benford's law. Stat Methods Appl. https://doi.org/10.1007/s10260-020-00547-1

FISCALIS (2016) The concept of tax gaps; report on VAT gap estimations. https://ec.europa.eu/taxation_customs/sites/taxation/files/docs/body/tgpg_report_en.pdf. FISCALIS Tax Gap Project Group FPG/041

Fuchs A, Letta G (1984) Sur le problème du premier chiffre décimal. Bollettino UMI 2(B):451–461

Fuchs A, Letta G (1996) Le problème du premier chiffre décimal pour les nombres premiers. Electron J Comb 3:R25

Giuliano Antonini R, Grekos G (2005) Regular sets and conditional density: an extension of Benford's law. Colloq Math 103:173–192

Gorroochurn P (2012) Classic problems of probability. Wiley, New York

Havil J (2008) Impossible? Surprising solutions to counterintuitive conundrums. Princeton University Press, Princeton

Herzel A (1956) Sulla distribuzione delle cifre iniziali dei numeri statistici. Atti della XV e XVI Riunione della Società Italiana di Statistica pp 205–228

Hill TP (1995a) The significant-digit phenomenon. Am Math Mon 102:322–327

Hill TP (1995b) A statistical derivation of the significant-digit law. Stat Sci 10:354–363

Knuth DE (1997) The art of computer programming, seminumerical algorithms, vol 2, 3rd edn. Addison-Wesley, Reading

Kossovsky AE (2015) Benford's law: theory, the general law of relative quantities, and forensic fraud detection applications. World Scientific, Singapore

Miller SJ (ed) (2015) Benford's law: theory and applications. Princeton Univ. Press, Princeton

Mumic N, Filzmoser P (2021) A multivariate test for detecting fraud based on Benford's law, with application to music streaming data. Stat Methods Appl. https://doi.org/10.1007/s10260-021-00582-6

Newcomb S (1881) Note on the frequency of use of the different digits in natural numbers. Am J Math 4:39–40

Nigrini MJ (1992) The detection of income tax evasion through an analysis of digital distributions. PhD thesis, Department of Accounting, University of Cincinnati

Nigrini MJ (2012) Benford's Law. Wiley, Hoboken

OLAF (2018) The OLAF report 2017. Eighteenth report of the European Anti-Fraud Office, 1 January to 31 December 2017. Tech. rep., European Anti-Fraud Office. https://doi.org/10.2784/93062

Olofsson L (2015) Probabilities: the little numbers that rule our lives, 2nd edn. Wiley, Hoboken

Perrotta D, Torti F (2010) Detecting price outliers in European trade data with the forward search. In: Palumbo F, Lauro C, Greenacre M (eds) Data analysis and classification. Studies in classification, data analysis, and knowledge organization. Springer, Berlin, Heidelberg

Perrotta D, Torti F (2018) Discussion of The power of monitoring: how to make the most of a contaminated multivariate sample. Stat Methods Appl 27:641–649

Perrotta D, Checchi E, Torti F, Cerasa A, Arnes Novau X (2020) Addressing price and weight heterogeneity and extreme outliers in surveillance data. Tech. Rep. JRC122315, European Commission, Joint Research Centre, Luxembourg. https://doi.org/10.2760/817681

Pickover C (2009) The math book. Sterling Publishing, New York

Regazzini E (1982) La legge di Benford–Furlan come legge statistica. Statistica 42:351–370

Riani M, Corbellini A, Atkinson AC (2018) The use of prior information in very robust regression for fraud detection. Int Stat Rev 86:205–218

Rousseeuw P, Perrotta D, Riani M, Hubert M (2019) Robust monitoring of time series with application to fraud detection. Econom Stat 9:108–121

Scala C (1966) Sulla stima statistica dell'evasione fiscale. G Econ Ann Econ 25(11/12):1198–1208

Scozzafava R (1981) Un esempio concreto di probabilità non-additiva: la distribuzione della prima cifra significativa dei dati statistici. Bollettino UMI 18(A):403–410

Stigler SM (1980) Stigler's law of eponymy. Trans N Y Acad Sci 39:147–157

Tijms H (2019) Surprises in probability. CRC Press, Boca Raton

Torti F, Riani M, Morelli G (2021) Semiautomatic robust regression clustering of international trade data. Stat Methods Appl. https://doi.org/10.1007/s10260-021-00569-3

Volčič A (1996) The first digit problem and scale invariance. In: Marcellini P, Talenti G, Vesentini E (eds) Partial differential equations and applications. Dekker, New York, pp 329–340

Wagon S (2010) Mathematica in action, 3rd edn. Springer, New York

Weyl H (1916) Über die gleichverteilung von zahlen mod eins. Math Ann 77:313–352