Benford's Law: Theory and Applications

Steven J Miller
Williams College

Benford's Law: Theory and Applications

Benford's Law: Theory and Applications

Edited by Steven J. Miller

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 6 Oxford Street, Woodstock, Oxfordshire OX20 1TW

press.princeton.edu

ISBN 978-0-691-14761-1

British Library Cataloging-in-Publication Data is available

This book has been composed in Times Roman in LaTeX

The publisher would like to acknowledge the authors of this volume for providing the camera-ready copy from which this book was printed.

Printed on acid-free paper. ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Dedicated to my colleagues and students for many fun years exploring Benford's Law together, to my parents Arlene and William for their support and encouragement over the years, and to the number 1 for being such a good, frequent companion. — SJM, Williamstown, MA, 2015.

The following is an excerpt from the book `Benford's Law: Theory and Applications`, published by Princeton University Press in 2015. For more information on the book, please see the publisher's homepage at `http://press.princeton.edu/titles/10527.html`.

Contents

Foreword

Perhaps I should immediately explain to the reader that this foreword is neither a voice from the grave, nor a remarkable display of anticipation. The "Frank A. Benford" who's the author of this foreword is a grandson of the "Frank A. Benford" for whom Benford's Law is named. As this relationship by itself hardly qualifies me to write this foreword, let me hasten to add that I'm a professional applied mathematician with a Ph.D. from Harvard.

That my grandfather's "Law of Anomalous Numbers" became known as "Benford's Law" instead of "Newcomb's Law" is, of course, a historical accident. I'm not complaining, obviously, but descendants of Simon Newcomb have a legitimate beef. I recently learned that my possibly distant cousin Gregory Benford, the well known physicist and science fiction author, had a colleague William Newcomb who was Simon Newcomb's grandson. Although Gregory and William worked closely together, the Benford/Newcomb connection never seems to have come up! Well, they're physicists, not mathematicians, so maybe I shouldn't be too surprised.

Figure 1 Frank Benford and his family, 1946.

I'd like to be able to claim that I remember grandfather, but that would not be true; he died when I was only three. I know that I met him, however, because a photograph (see Figure 1) from 1946 shows him and grandmother, their four sons, four daughters-in-law, and eight grandchildren at a family reunion. My mother was impressed by grandfather's ability to work calculus problems amid the hubbub of a family gathering.

My grandfather worked as a physicist in the Research Laboratory of the General Electric Company in Schenectady, NY, and most of his work for G.E. concerned optics. I'm sure that grandfather, being a conscientious employee, did all the research and writing of "The Law of Anomalous Numbers" on his own time. In a three-page autobiographical sketch he wrote for Leonard Clark of Union College in 1939, only one short paragraph concerns his Law of Anomalous Numbers.

Steven Miller, the editor of this volume, suggested that I include "stories about your grandfather, anything you know about the reception of his work, how he felt about it, how he would feel to see his name attached to something arising in so many different fields." As I don't have any first-hand information about grandfather, I passed this request along to my father, who wrote in his reply,

> My father was extremely modest and had little to say about his publications. He certainly never boasted. He was, indeed, interested in the phenomenon of first digits, but not excessively so. He would truly be surprised to learn of the interest that seems currently alive.

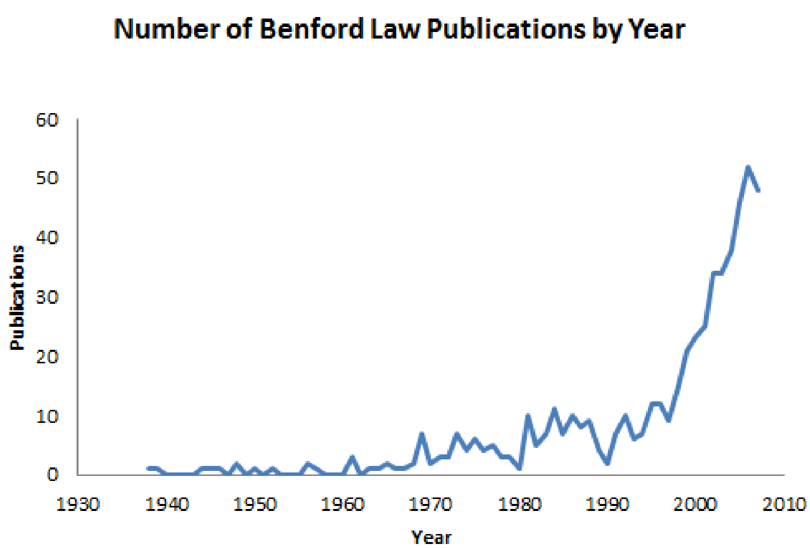


Figure 2 Publications on Benford's Law from 1938 to 2007.

I think my grandfather was proud of his paper, and fond of it, but he wasn't the sort to brag, and I suppose he was resigned to the idea that his Law of Anomalous

Numbers would soon disappear into obscurity. He would certainly be gratified (and possibly astonished) that the first-digit phenomenon he'd rediscovered seems to be a topic of perennial interest. As Ken Ross puts it, writing papers about Benford's Law appears to be "a growth industry." I offer as evidence some data I've compiled from the Benford Online Bibliography [BerH2]. Figure 2 shows the number of known "Benford relevant" publications per year for the first 70 years: 1938 until 2007; the last few years are omitted as there is a delay between when some papers are published and when they are added to the online bibliography. In Table 1 we give the frequencies of the first digits of the publication data from Figure 2 (omitting years where no relevant papers were published). The pattern of first digits should look familiar![1]

First digit	Frequency	Percentage	Benford's Law
1	20	35.7	30.1
2	9	16.1	17.6
3	8	14.3	12.5
4	5	8.9	9.7
5	3	5.4	7.9
6	2	3.6	6.7
7	6	10.7	5.8
8	1	1.8	5.1
9	2	3.6	4.6

Table 1 Frequencies in publication data of papers on Benford's Law: 1938 to 2007.

Roughly speaking, publications dealing with Benford's Law may be sorted into three classes: theoretical (extensions of Benford's Law, and investigations into the circumstances where Benford's Law does, or does not, apply), applications of Benford's Law, and popularizations (i.e., expository pieces aimed at the "intelligent layman"). I expect that almost all papers published between 1940 and 1965 are theoretical in nature. Starting with Ralph Raimi's 1969 Scientific American article [Rai], popular accounts of Benford's Law have appeared at a steady rate. This is attributable, of course, to the counterintuitive nature of the phenomenon. While theoretical papers on the first digit phenomenon have continued to appear, the publication of Mark Nigrini's 1992 dissertation [Nig1] marks the beginning of a wave of publications concerned with applications of Benford's Law. This is reflected, appropriately, in the contents of this book. There are 5 chapters that are theoretical in nature, and 13 chapters concerned with applications in accounting, vote fraud, economics, psychology, the natural sciences, and image processing.

It's been 75 years since my grandfather published "The Law of Anomalous Numbers," and it seems like a propitious time to publish a summary of the current state

[1]The chi-square statistic (comparing to the Benford frequencies) is 5.8 with 8 degrees of freedom (the 5% threshold value is 15.5, and thus the data is consistent with Benford behavior).

of affairs, i.e., the book you're reading. In the 75 years since grandfather wrote his article, the first-digit phenomenon has gone from an obscure curiosity to a fairly well-known and useful "law." Who knows what another 75 years will bring? There may be departments of Pure and Applied Benfordology at most major universities!

Frank A. Benford
BenfordAppliedMath.com
Salem, Oregon
June, 2014

Preface

One of the greatest beauties in mathematics is how the same equations can describe phenomena in widely different fields. Benford's Law of digit bias is an outstanding example of this. Briefly, it asserts that for many natural data sets we are more likely to see numbers with small leading digits than large ones. More precisely, our system is Benford if the probability of a first digit of d is $\log_{10} \frac{d+1}{d}$; we often consider the related but stronger statement that the probability of a significand being at most s is $\log_{10} s$, or the natural generalizations to other number bases. Base 10, the probabilities range from having a leading digit of 1 almost 30% of the time, to only about a 4.6% chance of starting with a 9.

Benford's Law arises in a variety of disciplines, including accounting, computer science, dynamical systems, economics, engineering, medicine, number theory, probability, psychology and statistics, to name just a few, and provides a wonderful opportunity for a common meeting ground for people with diverse interests and backgrounds. My first encounter with it was in Serre's *A Course in Arithmetic* [Ser]. On page 76 he remarks that Bombieri showed him a proof that the analytic density of the set of primes with leading digit 1 is $\log_{10} 2$, which is the Benford probability; a short argument using Poisson Summation yields the proof. I next saw it in Knuth's *The Art of Computer Programming, Volume 2: Seminumerical Algorithms* ([Knu], page 255), where he discusses applications of Benford's Law to analyzing floating point operations, especially the fact that Benford behavior implies the relative error from rounding is typically higher than one would expect. Once aware (or perhaps I should say doubly aware) of its existence, I saw it more and more often.

Our purposes here are to show students and researchers useful techniques from a variety of subjects, highlight the connections between the different areas and encourage research and cross-departmental collaboration on these problems. To do this, we develop much of the general theory in the first few chapters (concentrating on the methods which are applicable to a variety of problems), and then conclude with numerous chapters on applications written by world-experts in that field. Though there are common themes and methods throughout the applications, these chapters are self-contained, needing only the introductory chapters and some standard material. **For those wishing to use this as a textbook, numerous exercises and supplemental material are collected in the final chapter, and additionally are posted online (where more problems can easily be added, and links to relevant material for that chapter are collected); see** `http://web.williams.edu/Mathematics/sjmiller/public_html/benford/.`

One advantage of posting problems online is that this need not be a static list, and thus please feel free to email suggestions for additional exercises.

Below we briefly outline the major themes of the book.

- **Part I: General Theory I: Basis of Benford's Law:** We begin our study of Benford's Law with a brief introduction by Miller in Chapter 1. We concentrate on the history and some possible explanations, and briefly discuss a few of the many applications and central questions in the field.

  While for many readers this level of depth suffices, the subject can (and should!) be built on firm foundations. We do this in Chapter 2, where Berger and Hill rigorously derive many results through the use of appropriate σ-algebras. There are many approaches to proving a system satisfies Benford's Law. One of the most important is the Fundamental Equivalence (also called the uniform distribution characterization), which says a system $\{x_n\}$ satisfies Benford's Law base B if and only if its logarithm modulo 1 (i.e., $y_n = \log_B x_n \bmod 1$) is uniformly distributed. In other words, in the limit, the probability the logarithm modulo 1 lies in a subinterval $[a, b]$ of $[0, 1]$ is just $b - a$. The authors describe this and additional characterizations of Benford's Law (including the scale-invariance characterization and the base-invariance characterization), and prove many deterministic and random processes satisfy Benford's Law, as well as discussing flaws of other proposed explanations (such as the spread distribution approach).

  For the uniform distribution characterization to be useful, however, we need ways to show these logarithms are uniformly distributed. Often techniques from Fourier analysis are well suited for such an analysis. The Fundamental Equivalence reduces the Benfordness of $\{x_n\}$ to the distribution of the fractional parts of its logarithms $\{y_n\}$. Fourier analysis is built on the functions $e_m(t) := \exp(2\pi imt)$ (where $i = \sqrt{-1}$); note that the painful modulo condition in y_n vanishes when it is the argument of e_m, as $e_m(y_n) = e_m(y_n \bmod 1)$. Chapter 3 by Miller is devoted to developing Fourier analytic techniques to prove Benford behavior. We demonstrate the power of this machinery by applying it to a variety of problems, including products and chains of random variables, L-functions, special densities and the infamous $3x + 1$ problem. For example, using techniques from Fourier analysis (especially Poisson Summation), one can show that the standard exponential random variable is very close to satisfying Benford's Law. The exponential is a special case of the three-parameter Weibull distribution. A similar analysis shows that, so long as the shape exponent of the Weibull is not too large, it too is close to being Benford. There are numerous applications of these results. The closeness of the standard exponential to Benford implies that order statistics are almost Benford as well. The Weibull distribution arises in many survival models, and thus the analysis here provides another explanation of the prevalence of Benford behavior in many diverse systems.

- **Part II: General Theory II: Distributions and Rates of Convergence:** Combinations of data sets or random variables are often closer to satisfying Benford's Law than the individual data sets or distributions. This suggests a natural problem: looking for distributions that are exactly or at least close to being Benford. One of the most important examples of a distribution that exhibits Benford behavior is that of a geometric random variable. Numerous phenomena obey a geometric growth law; in particular, the solution to almost any linear difference equations is a linear combination of geometric series. We then investigate other important distributions and see how close they are to Benford. Although Benford's Law applies to a wide variety of data sets, none of the popular parametric distributions, such as the exponential and normal distributions, conforms exactly. Chapter 4 highlights the failure of several well-known probability distributions, then delves into the geometry associated with probability distributions that obey Benford's Law exactly. The starting point of these constructions is the fact that if U is a uniform random variable on $[a, a+n]$ for some integer n, then $T = 10^U$ is Benford base 10.

  As the exponential and Weibull distributions are not exactly Benford, it is important to obtain estimates on the size of the deviations. There are many ways to obtain such bounds. In Chapter 3 these bounds were obtained from Poisson Summation and the Fourier transform; in Chapter 5 Dümbgen and Leuenberger derive bounds from the total variation of the density (and its derivatives). These results are applied to numerous distributions, such as exponential, normal and Weibull random variables.

  This part concludes with Chapter 6 by Schürger. Earlier in the book we showed geometric Brownian motions are Benford. While processes such as the stock market were initially modeled by Brownian motions, such models have several defects, and current work must incorporate jumps and heavy tails. This leads to the study of Lévy Processes. These processes are described in detail, and their convergence to Benford behavior is shown. The techniques required are similar to those for geometric Brownian motion. On the other hand, the class of Lévy processes is much more general than just geometric Brownian motion, with applications in stochastic processes and finance; in particular, these and related processes model financial data, which has long been known to closely follow Benford's Law.

The final parts of this book deal with just some of the many applications of Benford's Law. Due to space constraints it is impossible to discuss all of the places Benford's Law appears. We have therefore chosen to focus on just a few situations, going for depth over breadth. We encourage the reader to peruse the many resources, such as the searchable online bibliography at [BerH2] or the large compilation [Hu], for a tour through additional areas to explore.

- **Part III: Applications I: Accounting and Vote Fraud:** Though initially an amusing observation about the distribution of digits in various data sets, since then Benford's Law has found numerous applications in many diverse fields. We briefly survey some of these. Probably the most famous application is to detecting tax fraud, though of course it is fruitfully used elsewhere too. We start in Chapter 7 with some of the basics of accounting, where Cleary and Thibodeau describe how Benford's Law can be integrated into business statistics and accounting courses. In particular, in the American Statistical Association's 2005 report *Guidelines for Assessment and Instruction in Statistics Education*, the following four goals (among others) are listed for what students should know after a first statistics course: (1) that variability is natural, predictable and quantifiable; (2) that random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken; (3) how to interpret statistical results in context; (4) how to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information. The rest of the chapter shows how incorporating Benford's Law realizes these objectives.

  Chapter 8 by Nigrini describes one of the most important applications of Benford's Law: detecting fraud. Many diverse systems approximately obey the law, and thus deviations often indicate fraud. The chapter begins by examining some data sets that follow the law (tax returns, the 2000 census, stream flow data and accounts payable data), and concludes by showing how Benford's Law successfully detected fraud in accounts payable amounts, payroll data and corporate numbers (such as Enron).

  We continue with another important example where Benford's Law has successfully detected fraud. Chapters 9 by Mebane and 10 by Roukema discuss how Benford's Law can detect vote fraud; the first chapter develops tests based on the second digit and explores its use in practice, while the second concentrates on a recent Iranian election whose official vote counts were claimed to be invalid. .

- **Part IV: Applications II: Economics:** While there is no dearth of interesting topics to explore, we have chosen to devote this part of the book to economics because of the huge impact of recent events. A spectacular example of this is given by European Union (EU) policy, and the situation in Greece. We begin in Chapter 11 by Rauch, Göttsche, Brähler and Engel with a description of EU practices and data from several countries. As the stakes are high, there is enormous pressure to misreport statistics to avoid being hit with EU deficit procedures. We continue in Chapter 12 by Tödter with additional analysis, especially of published economics research papers. A surprisingly large proportion of first digits of regression coefficients and standard errors violate Benford's Law, in contrast to second digits. Routine applications of Benford tests would increase the efficiency of replication exercises and raise the risk of scientific misconduct. Another issue discussed is fitting data to a Generalized Benford Law, a topic Lee, Cho and Judge address in Chapter

17; both of these chapters deal with the issues facing the public arising from researchers falsifying data. We conclude this part with an analysis of data from the U.S. financial sector. The main finding is that Benford's Law fits the data from before the housing crisis well, but not the data afterwards.

- **Part V: Applications III: Sciences:** In previous chapters we discussed which distributions fit (and which don't fit) Benford's Law, as well as tests to detect fraud. In this part we take a different approach, and explore the psychology behind the people generating numbers. Chapters 14 by Burns and Krygier and 15 by Chou, Kong, Teo and Zheng explore patterns and tendencies in number generation, and the resulting implications, followed by Hoyle's chapter on the prevalence of Benford's Law in the natural sciences, including a summary of its occurrences and a discussion of the consequences. We end in Chapter 17 by Lee, Cho and Judge with a nice mix of theory and application. The authors consider a generalization of Benford's Law, developing the theory and analyzing known cases of fraud. They study the related Stigler distribution, and describe how it may be found from information-theoretic methods. This leads to alternative digit distributions based on maximum entropy principles. The chapter ends by using these new distributions in an analysis of some medical data which was known to be falsified, where the falsified data is detected. An important application of the material of this part is in developing tests to detect whether researchers are submitting fraudulent data. Similar to the chapters from economics, as the costs to society from incorrectly adopting conclusions of faulty research can be high, these tests provide a valuable tool to check the veracity of claims.

- **Part VI: Applications IV: Images:** Our final part deals with whether or not images follow Benford's Law. Chiverton and Wells, in Chapter 18, explore the relationship between intensities in medical images and Benford behavior. They describe a simple classifier based on Bayes theory which uses the Benford Partial Volume (PV) distribution as a prior; the results show experimentally that the Benford PV distribution is a reasonable modeling tool for the classification of imaging data affected by the PV artifact. The fraud-based applications of Benford's Law have grown from financial data sets to others as well. The last chapter, Chapter 19 by Pérez-González, Quach, Abdallah, Heileman and Miller, explores whether or not Benford's Law can detect modifications in images. Specifically, while images in the pixel domain are not close to Benford, the result after applying the Discrete Cosine Transform is. These results can be used to look for hidden messages in pictures, as well as to test whether or not the image has been compressed.

We are extremely grateful to Princeton University Press, especially to our editor Vickie Kearn and to Betsy Blumenthal and Jill Harris, for all their help and aid, to our copyeditor Alison Durham who did a terrific job, especially in standardizing the exposition across chapters, to Meghan Kanabay for assistance with many of the illustrations, and Amanda Weiss for help with the jacket design. Many people proofread the book, looking not just for grammatical issues but also making sure it was a coherent whole with widely accessible expositions; it is a pleasure to thank them, especially John Bihn and Jaclyn Porfilio.

The editor was partially supported by NSF Grants DMS0600848, DMS0970067 and DMS1265673; some of his students assisting with the project were supported by NSF Grants DMS0850577 and DMS1347804, the Clare Boothe Luce Program, and Williams College. Some of this book is based on a conference organized by Chaouki T. Abdallah, Gregory L. Heileman, Steven J. Miller and Fernando Pérez-González and assisted by Ted Hill: *Conference on the Theory and Applications of Benford's Law* (16–18 December 2007, Santa Fe, NM). This conference was supported in part by Brown University, IEEE, NSF Grant DMS-0753043, the New Mexico Consortium's Institute for Advanced Study, Universidade de Vigo and the University of New Mexico, and it is a pleasure to thank them and the participants.

Steven J. Miller
Williams College
Williamstown, MA
October 2013

sjm1@williams.edu, Steven.Miller.MC.96@aya.yale.edu

Notation

$\square$: indicates the end of a proof.

$\equiv$: $x \equiv y \mod n$ means there exists an integer a such that $x = y + an$.

$\exists$: there exists.

$\forall$: for all.

$|\cdot|$: $|S|$ (or $\#S$) is the number of elements in the set S.

$\lceil\cdot\rceil$: $\lceil x\rceil$ is the smallest integer greater than or equal to x, read "the ceiling of x."

$\lfloor\cdot\rfloor$ or $[\cdot]$: $\lfloor x\rfloor$ (also written $[x]$) is the greatest integer less than or equal to x, read "the floor of x."

$\{\cdot\}$ or $\langle\cdot\rangle$: $\{x\}$ is the fractional part of x; note $x = [x] + \{x\}$.

$\ll, \gg$: see big-Oh notation.

$\vee$: $a \vee b$ is the maximum of a and b.

$\wedge$: $a \wedge b$ is the minimum of a and b.

$\mathbb{1}_A$ (or I_A) : the indicator function of set A; thus $\mathbb{1}_A(x)$ is 1 if $x \in A$ and 0 otherwise.

δ_a : Dirac probability measure concentrated at $a \in \Omega$.

λ : Lebesgue measure on $(\mathbb{R}, \mathcal{B})$ or parts thereof.

$\lambda_{a,b}$: normalized Lebesgue measure (uniform distribution) on $\big([a,b), \mathcal{B}[a,b)\big)$.

$\sigma(f)$: the σ-algebra generated by the function $f : \Omega \to \mathbb{R}$.

$\sigma(A)$: the spectrum (set of eigenvalues) of a $d \times d$-matrix A.

A^c : the complement of A in some ambient space Ω clear from the context; i.e., $A^c = \{\omega \in \Omega : \omega \notin A\}$.

$A \backslash B$: the set of elements of A not in B, i.e., $A \backslash B = A \cap B^c$.

$A \Delta B$: the symmetric difference of A and B, i.e., $A \Delta B = A \backslash B \cup B \backslash A$.

a.e. : (Lebesgue) almost every (or almost everywhere).

a.s. : almost surely, i.e., with probability one.

$\mathbb{B}$: Benford distribution on $(\mathbb{R}^+, \mathcal{S})$.

$\mathcal{B}$: Borel σ-algebra on $\mathbb{R}$ or parts thereof.

Big-Oh notation : $A(x) = O(B(x))$, read "$A(x)$ is of order (or big-Oh) $B(x)$," means there exists a $C > 0$ and an x_0 such that for all $x \geq x_0$, $|A(x)| \leq CB(x)$. This is also written $A(x) \ll B(x)$ or $B(x) \gg A(x)$.

$\mathbb{C}$: the set of complex numbers: $\{z : z = x + iy,\ x, y \in \mathbb{R}\}$.

C^ℓ : the set of all ℓ times continuously differentiable functions, $\ell \in \mathbb{N}_0$.

C^∞ : the set of all smooth (i.e., infinitely differentiable) functions; $C^\infty = \bigcap_{\ell \geq 0} C^\ell$.

$D_1, D_2, D_3, \ldots$: the first, second, third, $\ldots$ significant decimal digit.

$D_m^{(b)}$: the mth significant digit base b.

$\mathbb{E}[X]$ (or $\mathbb{E}X$) : the expectation of X.

$e(x)$: $e(x) = e^{2\pi i x}$.

$f_*\mathbb{P}$: a probability measure on $\mathbb{R}$ induced by $\mathbb{P}$ and the measurable function $f : \Omega \to \mathbb{R}$, via $f_*\mathbb{P}(\cdot) := \mathbb{P}\big(f^{-1}(\cdot)\big)$.

F_n: $\{F_n\}$ is the sequence of Fibonacci numbers, $\{F_n\} = \{0, 1, 1, 2, 3, 5, 8, \ldots\}$ ($F_{n+2} = F_{n+1} + F_n$ with $F_0 = 0$ and $F_1 = 1$).

F_P, F_X : the distribution functions of P and X.

i : $i = \sqrt{-1}$.

i.i.d. : independent, identically distributed (sequence or family of random variables); often one writes i.i.d.r.v.

$\Im z$: see $\Re z$.

infimum : the infimum of a set, denoted $\inf_n x_n$, is the largest number c (if one exists) such that $x_n \geq c$ for all n, and for any $\epsilon > 0$ there is some n_0 such that $x_{n_0} < c + \epsilon$. If the sequence has finitely many terms, the infimum is the same as the minimum value.

j : in some chapters $j = \sqrt{-1}$ (this convention is frequently used in engineering).

Leb : Lebesgue measure.

Little-Oh notation : $A(x) = o(B(x))$, read "$A(x)$ is little-Oh of $B(x)$," means $\lim_{x\to\infty} A(x)/B(x) = 0$.

$L^1(\mathbb{R})$: all $f : \mathbb{R} \longrightarrow \mathbb{C}$ which are measurable and Lebesgue integrable.

$\log$: usually the natural logarithm, though in some chapters it is the logarithm base 10.

$\ln$: the natural logarithm.

$\mathbb{N}$: the set of natural numbers: $\{0, 1, 2, 3, \dots\}$.

$\mathbb{N}_0$: the set of positive natural number: $\{1, 2, 3, \dots\}$.

N_f : the Newton map associated with a differentiable function f.

$o(\cdot)$, $O(\cdot)$: see "little-Oh" and "big-Oh" notation, respectively.

$O_T(x_0)$: the orbit of x_0 under the map T, possibly nonautonomous.

$\{p_n\}$: the set of prime numbers: 2, 3, 5, 7, 11, 13, $\dots$.

P : probability measure on $(\mathbb{R}, \mathcal{B})$, possibly random.

P_X : the distribution of the random variable X.

Prob (or Pr) : a probability function on a probability space.

$\mathbb{Q}$: the set of rational numbers: $\{x : x = \frac{p}{q}, p, q \in \mathbb{Z}, q \neq 0\}$.

$\mathbb{R}$: the set of real numbers.

$\mathbb{R}^+$: the set of positive real numbers.

$\Re z$, $\Im z$: the real and imaginary parts of $z \in \mathbb{C}$; if $z = x + iy$ then $\Re z = x$ and $\Im z = y$.

S : the significand function: if $x > 0$ then $x = S(x) \cdot 10^{k(x)}$, where $S(x) \in [1, 10)$ and $k(x) \in \mathbb{Z}$; more generally one can study the significand function S_B in base B.

$\mathcal{S}$: the significand σ-algebra.

supremum : given a sequence $\{x_n\}_{n=1}^{\infty}$, the supremum of the set, denoted $\sup_n x_n$, is the smallest number c (if one exists) such that $x_n \leq c$ for all n, and for any $\epsilon > 0$ there is some n_0 such that $x_{n_0} > c - \epsilon$. If the sequence has finitely many terms, the supremum is the same as the maximum value.

u.d. mod 1 : uniformly distributed modulo 1.

$\mathrm{Var}(X)$ (or $\mathrm{var}(X)$) : the variance of the random variable X, assuming the expected value of X is finite; $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

$\mathbb{W}$: the set of whole numbers: $\{1, 2, 3, 4, \dots\}$.

$X_n \stackrel{\mathcal{D}}{\to} X$: (X_n) converges in distribution to X.

$X_n \stackrel{\text{a.s.}}{\to} X$: (X_n) converges to X almost surely.

$\overline{z}$, $|z|$: the conjugate and absolute value of $z \in \mathbb{C}$.

$\mathbb{Z}$: the set of integers: $\{\dots, -2, -1, 0, 1, 2, \dots\}$.

$\mathbb{Z}^+$: the set of non-negative integers, $\{0, 1, 2, \dots\}$.

PART I
General Theory I: Basis of Benford's Law

Chapter One

A Quick Introduction to Benford's Law

Steven J. Miller[1]

The history of Benford's Law is a fascinating and unexpected story of the interplay between theory and applications. From its beginnings in understanding the distribution of digits in tables of logarithms, the subject has grown enormously. Currently hundreds of papers are being written by accountants, computer scientists, engineers, mathematicians, statisticians and many others. In this chapter we start by stating Benford's Law of digit bias and describing its history. We discuss its origins and give numerous examples of data sets that follow this law, as well as some that do not. From these examples we extract several explanations as to the prevalence of Benford's Law, which are described in greater detail later in the book. We end by quickly summarizing many of the diverse situations in which Benford's Law holds, and why an observation that began in looking at the wear and tear in tables of logarithms has become a major tool in subjects as diverse as detecting tax fraud and building efficient computers. We then continue in the next chapters with rigorous derivations, and then launch into a survey of some of the many applications. In particular, in the next chapter we put Benford's Law on a solid foundation. There we explore several different categorizations of Benford's Law, and rigorously prove that certain systems satisfy these conditions.

1.1 OVERVIEW

We live in an age when we are constantly bombarded with massive amounts of data. Satellites orbiting the Earth daily transmit more information than is in the entire Library of Congress; researchers must quickly sort through these data sets to find the relevant pieces. It is thus not surprising that people are interested in patterns in data. One of the more interesting, and initially surprising, is Benford's Law on the distribution of the first or the leading digits.

In this chapter we concentrate on a mostly non-technical introduction to the subject, saving the details for later. Before we can describe the law, we must first set notation. At some point in secondary school, we are introduced to **scientific notation**: any positive number x may be written as $S(x) \cdot 10^k$, where $S(x) \in [1, 10)$ is the **significand** and k is an integer (called the **exponent**). The integer part of the

[1]Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267. The author was partially supported by NSF grants DMS0970067 and DMS1265673.

significand is called the **leading digit** or the **first digit**. Some people prefer to call $S(x)$ the mantissa and not the significand; unfortunately this can lead to confusion, as the **mantissa** is the fractional part of the logarithm, and this quantity too will be important in our investigations. As always, examples help clarify the notation. The number 1701.24601 would be written as $1.70124601 \cdot 10^3$ in scientific notation. The significand is 1.70124601, the exponent is 3 and the leading digit is 1. If we take the logarithm base 10, we find $\log_{10} 1701.24601 \approx 3.2307671196444460726$, so the mantissa is approximately .2307671196444460726.

There are many advantages to studying the first digits of a data set. One reason is that it helps us compare apples and apples and not apples and oranges. By this we mean the following: two different data sets could have very different scales; one could be masses of subatomic particles while another could be closing stock prices. While the units are different and the magnitudes differ greatly, every number has a unique leading digit, and thus we can compare the distribution of the first digits of the two data sets.

The most natural guess would be to assert that for a generic data set, all numbers are equally likely to be the leading digit. We would then posit that we should observe about 11% of the time a leading digit of 1, 2, $\dots$, 9 (note that we would guess each number occurs one-ninth of the time and not one-tenth of the time, as 0 is the leading digit for only one number, namely 0). The content of Benford's Law is that this is frequently not so; specifically, in many situations we expect the leading digit to be d with probability approximately $\log_{10}\left(\frac{d+1}{d}\right)$, which means the probability of a first digit of 1 is about 30% while a first digit of 9 happens about 4.6% of the time.

1.2 NEWCOMB

Though it is called Benford's Law, he was not the first to observe this digit bias. Our story begins with the astronomer–mathematician Simon Newcomb, who observed this behavior more than 50 years before Benford. Newcomb was born in Nova Scotia in 1835 and died in Washington, DC in 1909. In 1881 he published a short article in the American Journal of Mathematics, *Note on the Frequency of Use of the Different Digits in Natural Numbers* (see [New]). The article begins,

> That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9. The question naturally arises whether the reverse would be true of logarithms. That is, in a table of anti-logarithms, would the last part be more used than the first, or would every part be used equally? The law of frequency in the one case may be deduced from that in the other. The question we have to consider is, what is the probability that if a natural number be taken at random its first significant digit will be n, its second n', etc.

> As natural numbers occur in nature, they are to be considered as the ratios of quantities. Therefore, instead of selecting a number at random, we must select two numbers, and inquire what is the probability that the first significant digit of their ratio is the digit n. To solve the problem we may form an indefinite number of such ratios, taken independently; and then must make the same inquiry respecting their quotients, and continue the process so as to find the limit towards which the probability approaches.

In this short article two very important properties of the distribution of digits are noted. The first is that all digits are not equally likely. The article ends with a quantification of how oftener the first digit is a 1 than a 9, with Newcomb stating,

> The law of probability of the occurrence of numbers is such that all mantissæ of their logarithms are equally probable.

Specifically, Newcomb gives a table (see Table 1.1) for the probabilities of first and second digits.

d	Probability first digit d	Probability second digit d
0		0.1197
1	0.3010	0.1139
2	0.1761	0.1088
3	0.1249	0.1043
4	0.0969	0.1003
5	0.0792	0.0967
6	0.0669	0.0934
7	0.0580	0.0904
8	0.0512	0.0876
9	0.0458	0.0850

Table 1.1 Newcomb's conjecture for the probabilities of observing a first digit of d or a second digit of d; all probabilities are reported to four decimal digits.

The second key observation of his paper is noting the importance of scale. The numerical value of a physical quantity clearly depends on the scale used, and thus Newcomb suggests that the correct items to study are ratios of measurements.

1.3 BENFORD

The next step forward in studying the distribution of the leading digits of numbers was Frank Benford's *The Law of Anomalous Numbers*, published in the Proceedings of the American Philosophical Society in 1938 (see [Ben]). In addition to advancing explanations as to why digits have this distribution, he also presents some justification as to why this is a problem worthy of study.

> It has been observed that the pages of a much used table of common logarithms show evidences of a selective use of the natural numbers. The pages containing the logarithms of the low numbers 1 and 2 are apt to be more stained and frayed by use than those of the higher numbers 8 and 9. Of course, no one could be expected to be greatly interested in the condition of a table of logarithms, but the matter may be considered more worthy of study when we recall that the table is used in the building up of our scientific, engineering, and general factual literature. There may be, in the relative cleanliness of the pages of a logarithm table, data on how we think and how we react when dealing with things that can be described by means of numbers.

Benford studied the distribution of leading digits of 20 sets of data, including rivers, areas, populations, physical constants, mathematical sequences (such as $\sqrt{n}$, $n!$, n^2, $\dots$), sports, an issue of Reader's Digest and the first 342 street addresses given in the (then) current American Men of Science. We reproduce his observations in Table 1.2.

Title	1	2	3	4	5	6	7	8	9	Count
Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
$n^{-1}, \sqrt{n}$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
Digest	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
$n, n^2, \dots, n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Benford's Law	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6	

Table 1.2 Distribution of leading digits from the data sets of Benford's paper [Ben]; the amalgamation of all observations is denoted by "Average." Note that the agreement with Benford's Law is better for some examples than others, and the amalgamation of all examples is fairly close to Benford's Law.

Benford's paper contains many of the key observations in the subject. One of the most important is that while individual data sets may fail to satisfy Benford's Law, amalgamating many different sets of data leads to a new sequence whose behavior is typically closer to Benford's Law. This is seen both in the row corresponding to $n, n^2, \dots$ (where we can prove that each of these is non-Benford) as well as in the average over all data sets.

Benford's article suffered a much better fate than Newcomb's paper, possibly in part because it immediately preceded a physics article by Bethe, Rose and Smith on the multiple scattering of electrons. Whereas it was decades before there was another article building on Newcomb's work, the next article after Benford's paper was six years later (by S. A. Goutsmit and W. H. Furry, *Significant Figures of Numbers in Statistical Tables*, in Nature), and after that the papers started occurring more and more frequently. See Hurlimann's extensive bibliography [Hu] for a list of papers, books and reports on Benford's Law from 1881 to 2006, as well as the online bibliography maintained by Arno Berger and Ted Hill [BerH2].

1.4 STATEMENT OF BENFORD'S LAW

We are now ready to give precise statements of Benford's Law.

Definition 1.4.1 (**Benford's Law for the Leading Digit**). *A set of numbers satisfies Benford's Law for the Leading Digit if the probability of observing a first digit of* d *is* $\log_{10}\left(\frac{d+1}{d}\right)$.

While clean and easy to state, the above definition has several problems when we apply it to real data sets. The most glaring is that the numbers $\log_{10}\left(\frac{d+1}{d}\right)$ are irrational. If we have a data set with N observations, then the number of times the first digit is d must be an integer, and hence the observed frequencies are always rational numbers.

One solution to this issue is to consider only infinite sets. Unfortunately this is not possible in many cases of interest, as most real-world data sets are finite (i.e., there are only finitely many counties or finitely many trading days). Thus, while Definition 1.4.1 is fine for mathematical investigations of sequences and functions, it is not practical for many sets of interest. We therefore adjust the definition to

Definition 1.4.2 (**Benford's Law for the Leading Digit (Working Definition)**). *We say a data set satisfies Benford's Law for the Leading Digit if the probability of observing a first digit of* d *is approximately* $\log_{10}\left(\frac{d+1}{d}\right)$.

Note that the above definition is vague, as we need to clarify what is meant by "approximately." It is a non-trivial task to find good statistical tests for large data sets. The famous and popular chi-square tests, for example, frequently cannot be used with extensive data sets as this test becomes very sensitive to small deviations when there are many observations. For now, we shall use the above definition and interpret "approximately" to mean a good visual fit. This approach works quite well for many applications. For example, in Chapter 8 we shall see that many corporate and other financial data sets follow Benford's Law, and thus if the distribution is visually far from Benford, it is quite likely that the data's integrity has been compromised.

Finally, instead of studying just the leading digit we could study the entire significand. Thus in place of asking for the probability of a first digit of 1 or 2 or 3, we now ask for the probability of observing a significand between 1 and 2, or between π and e. This generalization is frequently called the **Strong Benford's Law**.

Definition 1.4.3 (**Strong Benford's Law for the Leading Digits (Working Definition)).** *We say a data set satisfies the Strong Benford's Law if the probability of observing a significand in* $[1, s)$ *is* $\log_{10} s$.

Note that Strong Benford behavior implies Benford behavior; the probability of a first digit of d is just the probability the significand is in $[d, d+1)$. Writing $[d, d+1)$ as $[1, d+1)\backslash[1, d)$, we see this probability is just $\log_{10}(d+1)-\log_{10} d = \log_{10} \frac{d+1}{d}$.

1.5 EXAMPLES AND EXPLANATIONS

In this section we briefly give some explanations for why so many different and diverse data sets satisfy Benford's Law, saving for later chapters more detailed explanation. It's worthwhile to take a few minutes to reflect on how Benford's Law was discovered, and to see whether or not similar behavior might be lurking in other systems. The story is that Newcomb was led to the law by observing that the pages in logarithm tables corresponding to numbers beginning with 1 were significantly more worn than the pages corresponding to numbers with higher first digit. A reasonable explanation for the additional wear and tear is that numbers with a low first digit are more common than those with a higher first digit. It is thus quite fortunate for the field that there were no calculators back then, as otherwise the law could easily have been missed. Though few (if any) of us still use logarithm tables, it is possible to see a similar phenomenon in the real world today. Our analysis of this leads to one of the most important theorems in probability and statistics, the Central Limit Theorem, which plays a role in understanding the ubiquity of Benford's Law.

Instead of looking at logarithm tables, we can look at the steps in an old building, or how worn the grass is on college campuses. Assuming the steps haven't been replaced and that there is a reasonable amount of traffic in and out of the building, then lots of people will walk up and down these stairs. Each person causes a small amount of wear and tear on the steps; though each person's contribution is small, if there are enough people over a long enough time period then the cumulative effect will be visually apparent. Typically the steps are significantly more worn towards the center and less so as one moves towards the edges. A little thought suggests the obvious answer: people typically walk up the middle of a flight of stairs unless someone else is coming down. Similar to carbon dating, one could attempt to determine the age of a building by the indentation of the steps. Looking at these patterns, we would probably see something akin to the normal distribution, and if we were fortunate we might "discover" the Central Limit Theorem. There are many other examples from everyday life. We can also observe this in looking at lawns. Everyone knows the shortest distance between two points is a line, and people frequently leave the sidewalks and paths and cut across the grass, wearing

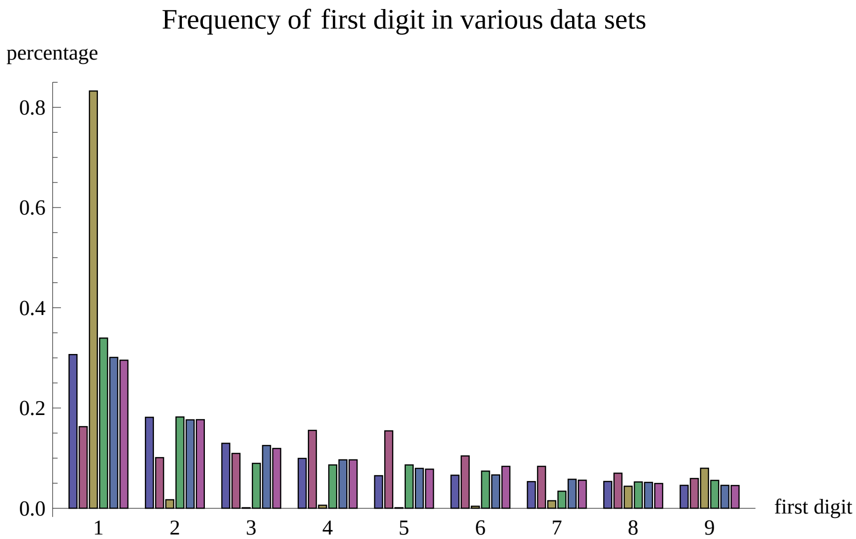


Figure 1.1 Frequencies of leading digits for (a) U.S. county populations (from 2000 census); (b) U.S. county land areas in square miles (from 2000 census); (c) daily volume of NYSE trades from 2000 through 2003; (d) fundamental constants (from NIST); (e) first 3219 Fibonacci numbers; (f) first 3219 factorials. Note the census data includes Puerto Rico and the District of Columbia.

it down to dirt in some places and leaving it untouched in others. Another example is to look at keyboards, and compare the well-worn "E" to the almost pristine "Q." Or the wear and tear on doors. The list is virtually endless.

In Figure 1.1 we look at the leading digits of the several "natural" data sets. Four arise from the real world, coming from the 2000 census in the United States (population and area in square miles of U.S. counties), daily volumes of transactions on the New York Stock Exchange (NYSE) from 2000 through 2003 and the physical constants posted on the homepage of the National Institute for Standards and Technology (NIST); the remaining two data sets are popular mathematical sequences: the first 3219 Fibonacci numbers and factorials (we chose this number so that we would have as many entries as we do counties).

If these are "generic" data sets, then we see that no one law describes the behavior of each set. Some of the sets are quite close to following Benford's Law, others are far off; none are close to having each digit equally likely to be the leading digit. Except for the second and third sets, the rest of the data behaves similarly; this is easier to see if we remove these two examples, which we do in Figure 1.2.

Before launching into explanations of why so many data sets are Benford (or at least close to it), it's worth briefly remarking why many are not. There are several reasons and ways a data set can fail to be Benford; we quickly introduce some of these reasons now, and expand on them more when we advance explanations for Benford's Law below. For example, imagine we are recording hourly temperatures

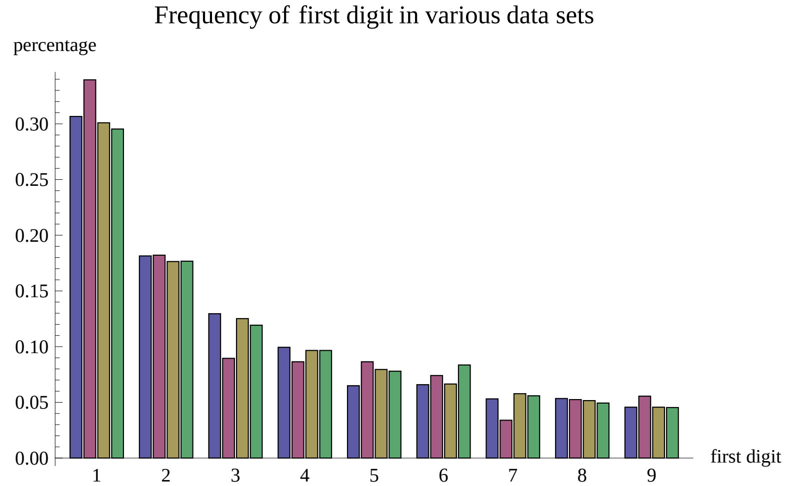


Figure 1.2 Frequencies of leading digits for (a) U.S. county populations (from 2000 census); (b) fundamental constants (from NIST); (c) first 3219 Fibonacci numbers; (d) first 3219 factorials. Note the census data includes Puerto Rico and the District of Columbia.

in May at London Heathrow Airport. In Fahrenheit the temperatures range from lows of around 40 degrees to highs of around 80. As all digits are not accessible, it's impossible to be Benford, though perhaps *given this restriction, the relative probabilities of the digits are Benford*.

For another issue, we have many phenomena that are given by specific, concentrated distributions that will not be Benford. The Central Limit Theorem is often a good approximation for the behavior of numerous processes, ranging from heights and weights of people to batting averages to scores on exams. In these situations we clearly do not expect Benford behavior, though we will see below that processes whose *logarithms* are normally distributed (with large standard deviations) are close to Benford.

Thus, in looking for data sets that are close to Benford, it is natural to concentrate on situations where the values are not given by a distribution concentrated in a small interval. We now explore some possibilities below.

1.5.1 The Spread Explanation

We drew the examples in Figure 1.1 from very different fields; why do so many of them behave similarly, and why do others violently differ? While the first question still confounds researchers, we can easily explain why two data sets had such different behavior, and this reason has been advanced by many as a source of Benford's Law (though there are issues with it, which we'll comment on shortly). Let's look at the first two sets of data: the population in U.S. counties in 2000 and daily volume of the NYSE from 2000 through 2003. You can see from the histogram in

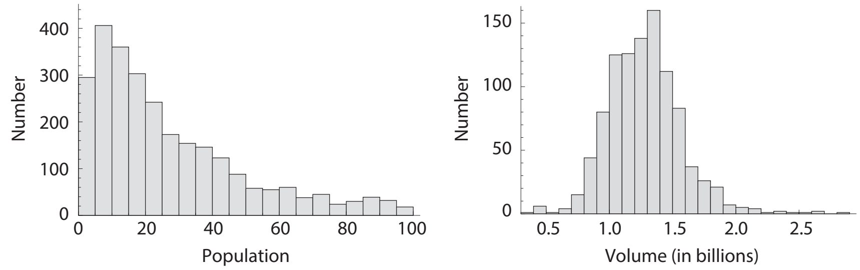


Figure 1.3 (Left) The population (in thousands) of U.S. counties under 250,000 (which is about 84% of all counties). (Right) The daily volume of the NYSE from 2000 through 2003. Note the population spans two orders of magnitude while the stock volumes are mostly within a factor of 2 of each other.

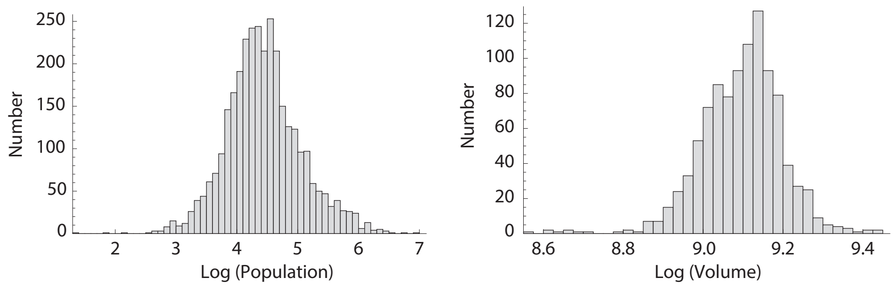


Figure 1.4 (Left) The population of U.S. counties. (Right) The daily volume of the NYSE from 2000 through 2003.

Figure 1.3 the stock market transactions are clustered around one value and span only one order of magnitude. Thus it is not surprising that there is little variation in these first digits. For the county populations, however, the data is far more spread out. These effects are clearer if we look at a histogram of the log-plot of the data, which we do in Figure 1.4. A detailed analysis of the other data sets shows similar behavior; the four data sets that behave similarly are spread out on a logarithmic plot over several orders of magnitude, while the two sets that exhibit different behavior are more clustered on a log-plot.

Our discussion above leads to our first explanation for Benford's Law, the **spread hypothesis**. The spread hypothesis states that if a data set is distributed over several orders of magnitude, then the leading digits will approximately follow Benford's Law. Of course, a little thought shows that we need to assume far more than the data just being spread out over several orders of magnitude. For example, if our set of observations were

$$\{1,\ 10,\ 100,\ 1000,\ \dots,\ 10^{2015}\}$$

then clearly it is non-Benford, even though it does cover over 2000 orders of magnitude! As remarked above, our purpose in this introduction is to just briefly introduce the various ideas and approaches, saving the details for later. There are many issues with the **spread hypothesis**; see Chapter 2 and [BerH3] for an excellent analysis of these problems.

1.5.2 The Geometric Explanation

Our next attempt to explain the prevalence of Benford's Law goes back to Benford's paper [Ben], whose second part is titled *Geometric Basis of the Law*. The idea is that if we have a process with a constant growth rate, then more time will be spent at lower digits than higher digits. For definiteness, imagine we have a stock that increases at 4% per year. The amount of time it takes to move from \$1 to \$2 is the same as it would take to move from \$10,000 to \$20,000 or from \$100,000,000 to \$200,000,000. If n_d is the number of years it takes to move from d dollars to $d+1$ dollars then $d \cdot (1.04)^{n_d} = (d+1)$, or

$$n_d \;=\; \frac{\log\left(\frac{d+1}{d}\right)}{\log 1.04}. \tag{1.1}$$

In Table 1.3 we consider the (happy) situation of a stock that rises 4% each and every year. Notice that it takes over 17 years to move from being worth \$1 to being worth \$2, but less than 3 years to move from being worth \$9 to \$10.

First digit	Years	Percentage of time	Benford's Law
1	17.6730	0.30103	0.30103
2	10.3380	0.17609	0.17609
3	7.3350	0.12494	0.12494
4	5.6894	0.09691	0.09691
5	4.6486	0.07918	0.07918
6	3.9303	0.06695	0.06695
7	3.4046	0.05799	0.05799
8	3.0031	0.05115	0.05115
9	2.6863	0.04576	0.04576

Table 1.3 How long the first digit of a stock has leading digit d, given that the stock rises 4% each year. It takes the stock approximately 58.7084 years to increase from \$1 to \$10.

A little algebra shows that this implies Benford behavior. If n is the amount of time it takes to move from \$1 to \$10, then $1 \cdot (1.04)^n = 10$ or $n = \frac{\log 10}{\log 1.04}$. Thus by (1.1), we see the percentage of the time spent with a first digit of d is

$$\frac{\log\left(\frac{d+1}{d}\right)}{\log 1.04} \Bigg/ \frac{\log 10}{\log 1.04} \;=\; \frac{\log\left(\frac{d+1}{d}\right)}{\log 10} \;=\; \log_{10}\left(\frac{d+1}{d}\right), \tag{1.2}$$

which is just Benford's Law! There is nothing special about 4%; the same analysis works in general *provided* that at each moment we grow by the same, fixed rate. The analysis is more interesting if at each instance the growth percentage is a random variable, say drawn from a Gaussian. For more on such processes see Chapter 6.

This is not an isolated example. Many natural and mathematical phenomena are governed by geometric growth. Examples range from radioactive decay and bacteria populations to the Fibonacci numbers. One reason for this is that solutions to many difference equations are given by linear combinations of geometric series; as difference equations are just discrete analogues of differential equations, it is thus not surprising that they model many situations. For example, the Fibonacci numbers satisfy the second order linear recurrence relation

$$F_{n+2} \;=\; F_{n+1} + F_n. \tag{1.3}$$

Once the first two Fibonacci numbers are known, the recurrence (1.3) determines the rest. If we start with $F_0 = 0$ and $F_1 = 1$, we find $F_2 = 1$, $F_3 = 2$, $F_4 = 3$, $F_5 = 5$ and so on. Moreover, there is an explicit formula for the nth term, namely

$$F_n \;=\; \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n; \tag{1.4}$$

known as Binet's formula; generalizations of it hold for solutions to linear recurrence relations. As $\left|\frac{1+\sqrt{5}}{2}\right| > 1$ and $\left|\frac{1-\sqrt{5}}{2}\right| < 1$, for large n this implies $F_n \approx \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n$. Note that $F_{n+1} \approx \frac{1+\sqrt{5}}{2}F_n$, or $F_{n+1} \approx 1.61803F_n$. This means that the Fibonacci numbers are well approximated by what would be a highly desirable stock rising about 61.803% each year, and hence by our previous analysis it is reasonable to expect the Fibonacci numbers will be Benford as well.

While the discreteness of the Fibonacci numbers makes the analysis a bit more complicated than the continuous growth rate problem, a generalization of these methods proves that the Fibonacci numbers, as well as the solution to many difference equations, are Benford. Again, our purpose here is to merely provide some evidence as to why so many different, diverse systems satisfy Benford's Law. It is not the case that every recurrence relation leads to Benford behavior. To see this, consider $a_{n+2} = 2a_{n+1} - a_n$ with either $a_0 = a_1 = 1$ (which implies $a_n = 1$ for all n) or $a_0 = 0$ and $a_1 = 1$ (which implies $a_n = n$ for all n). While there are examples of recurrence relations that are non-Benford, a "generic" one will satisfy Benford's Law, and thus studying these systems provides another path to Benford.

1.5.3 The Scale-Invariance Explanation

For our next explanation, we return to a comment from Newcomb's [New] paper:

> As natural numbers occur in nature, they are to be considered as the ratios of quantities. Therefore, instead of selecting a number at random, we must select two numbers, and inquire what is the probability that the first significant digit of their ratio is the digit n.

The import of this comment is that the behavior should be independent of the units used. For example, if we look at the value of stocks in our portfolio then the magnitudes will change if we measure their worth in dollars or euros or yen or bars of gold pressed latinum, though the physical quantities are unchanged. Similarly we can use the metric system or the (British) imperial system in measuring physical constants. As the universe doesn't care what units we use for our experiments, it is natural to expect that the distribution of leading digits should be unchanged if we change our units.

For definiteness, let's consider the areas of the countries in the world. There are almost 200 countries; if we measure area in square kilometers then about 28.49% have a first digit of 1 and 18.99% have a first digit of 2, while if we measure in square miles it is 34.08% have a first digit of 1 and 16.20% have a first digit of 2, which should be compared to the Benford probabilities of approximately 30.10% and 17.61%; one observes a similar closeness with the other digits.

The assumption that there *is* a distribution of the first digit and that this distribution is independent of scale implies the first digits follow Benford's Law. The analysis of this involves introducing a σ-algebra and studying scale-invariant probability measures on this space. Without going into these details now, we can at least show that Benford's Law is consistent with scale invariance.

Let's assume our data set satisfies the Strong Benford Law (see Definition 1.4.3). Then the probability the significand is in $[a, b] \subset [1, 10)$ is $\log_{10}(b/a)$. Assume now we rescale every number in our set by multiplying by a fixed constant C. For definiteness we take $C = \sqrt{3}$ and compute the probability that numbers in the scaled data set have leading digit 1. Note that multiplying $[1, 10)$ by $\sqrt{3}$ gives us the interval $[\sqrt{3}, 10\sqrt{3}) \approx [1.73, 17.32)$. The parts of this new interval with a leading digit of 1 are $[\sqrt{3}, 2)$ and $[10, 10\sqrt{3})$, which come from $[1, 2/\sqrt{3})$ and $[10/\sqrt{3}, 10)$. As we are assuming the strong form of Benford's Law, the probabilities of these two intervals are $\log_{10} \frac{2/\sqrt{3}}{1}$ and $\log_{10} \frac{10}{10\sqrt{3}}$. Summing yields the probability of the first digit of the scaled set being 1 is

$$\log_{10}\left(\frac{2/\sqrt{3}}{1}\right) + \log_{10}\left(\frac{10}{10\sqrt{3}}\right) \;=\; \log_{10} 2,$$

which is the Benford probability! A similar analysis works for the other leading digits and other choices of C.

We close this section by noting that scale invariance fits naturally with the other explanations introduced to date. If our initial data set were spread out over several orders of magnitude, so too would the scaled data. Similarly, if we return to our hypothetical stock increasing by 4% per year, the effect of changing the units of our currency can be viewed as changing our principal; however, what governs how long our stock spends with a leading digit of d is not the principal but rather the rate of growth, and that is unchanged.

1.5.4 The Central Limit Explanation

We need to introduce some machinery for our last heuristic explanation. If $y \geq 0$ is a real number, by $y \bmod 1$ we mean the fractional part of y. Other notations for this are $\{y\}$ or $y - \lfloor y \rfloor$. If $y < 0$ then $y \bmod 1$ is $1 - (-y \bmod 1)$. In other words, $y \bmod 1$ is the unique number in $[0, 1)$ such that $y - (y \bmod 1)$ is an integer. Thus $3.14 \bmod 1$ is .14, while $-3.14 \bmod 1$ is .86. We say y **modulo** 1 for $y \bmod 1$.

Recall that any positive number x may be written in **scientific notation** as $x = S(x) \cdot 10^k$, where $S(x) \in [1, 10)$ and k is an integer. The real number $S(x)$, called the **significand**, encodes all the information about the digits of x; the effect of k is to specify the decimal point's location. Thus, if we are interested in either the first digit or the significand, the value of k is immaterial. This suggests that rather than studying our data as given, it might be worthwhile to transform the data as follows:

$$x \ \mapsto \ \log_{10} x \bmod 1. \tag{1.5}$$

A little algebra shows that two positive numbers have the same leading digits if and only if their signficands have the same first digit. Thus if we have a set of values $\{x_1, x_2, x_3, \dots\}$ then the subset with leading digit d is $\{x_i : S(x_i) \in [d, d+1)\}$, which is equivalent to $\{x_i : \log_{10} S(x_i) \in [\log_{10} d, \log_{10}(d+1))\}$.

This innocent-looking reformulation turns out to be not only one of the most fruitful ways of exploring Benford's Law, but also highlights what is going on. We first explain the new perspective gained by transforming the data. According to Benford's Law, the probability of observing a first digit of d is $\log_{10} \frac{d+1}{d}$. This is $\log_{10}(d+1) - \log_{10} d$, which is the length of the interval $[\log_{10} d, \log_{10}(d+1))$! In other words, consider a data set satisfying Benford's Law, and transform the set as in (1.5). The new set lives in $[0, 1)$ and is uniformly distributed there. Specifically, the probability that we have a value in the interval $[\log_{10} d, \log_{10}(d+1))$ is the length of that interval.

While it may not seem natural to take the logarithm base 10 of each number, and then look at the result modulo 1, under such a process the resulting values are uniformly distributed if the initial set obeys Benford's Law. Another way of looking at this is that there is a natural transformation which takes a set satisfying Benford's Law and returns a new set of numbers that is uniformly distributed.

We briefly comment on why this is a natural process. We replace x with $\log_{10} x$ $\bmod 1$. If we write $x = S(x) \cdot 10^k$, then $\log_{10} x \bmod 1$ is just $\log_{10} S(x)$. Thus taking the logarithm modulo 1 is a way to get our hands on the significand (actually, its logarithm), which is what we want to understand. While the logarithm function is a nice function, removing the integer part *in general* is messy and leads to complications; however, there is a very important situation where it is painless to remove the integer part. Recall the exponential function

$$e(x) \ := \ e^{2\pi i x} \ = \ \cos(2\pi x) + i \sin(2\pi x), \tag{1.6}$$

where $i = \sqrt{-1}$. As $e(x+1) = e(x)$, we see

$$e(x \bmod 1) \ = \ e(x). \tag{1.7}$$

The utility of the above becomes apparent when we apply Fourier analysis. In Fourier analysis one uses sines, cosines or exponential functions to understand more complicated functions. From our analysis above, we may either include the modulo 1 or not in the argument of the exponential function. While we will elaborate on this at great length later, the key takeaway is that the transformed data is ideally suited for Fourier analysis.

We can now sketch how this is related to Benford's Law. There are many data sets in the world whose values are the product of numerous measurements. For example, the monetary value of a gold brick is a product of the brick's length, width, height, density and value of gold per pound. Imagine we have some quantity X which is a product of n values, so

$$X \ = \ X_1 \cdot X_2 \cdot \ldots \cdot X_n.$$

We assume the X_i's are nice random variables. From our discussion above, to show that X obeys Benford's Law it suffices to know that the distribution of the logarithm of X modulo 1 is uniformly distributed. Thus we are led to study

$$\log_{10} X \ = \ \log_{10}(X_1 \cdot X_2 \cdot \ldots \cdot X_n) \ = \ \log_{10} X_1 + \cdots + \log_{10} X_n.$$

By the Central Limit Theorem, if n is large then the above sum is approximately normally distributed, and the variance will grow with n; however, what we are really interested in is not this sum but rather this sum modulo 1:

$$\log_{10} X \bmod 1 \ = \ (\log_{10} X_1 + \cdots + \log_{10} X_n) \bmod 1.$$

A nice computation shows that as the variance σ tends to infinity, if we look at the probability density of a normal with variance σ modulo 1 then that is approximately uniformly distributed on $[0,1]$. Explicitly, let Y be normally distributed with some mean μ and very large variance σ. If we look at the probability density of the new random variable $Y \bmod 1$, then this is approximately uniformly distributed on $[0,1)$. This means that the probability that $Y \in [\log_{10} d, \log_{10}(d+1))$ is just $\log_{10}(d+1) - \log_{10} d$, or $\log_{10} \frac{d+1}{d}$; however, note that these are just the Benford probabilities!

While we have chosen to give the argument for multiplying random variables, similar results hold for other combinations (such as addition, exponentiation, etc.). The Central Limit Theorem is lurking in the background, and if we adjust our viewpoint we can see its effect.

1.6 QUESTIONS

Our goal in this book is to explain the prevalence of Benford's Law, and discuss its implications and applications. The question of leading digits is but one of many that we could ask. There are many generalizations; below we state the two most common.

1. *Instead of studying the distribution of the first digit, we may study the distribution of the first two, three, or more generally the significand, of our number. The Strong Benford's Law is that the probability of observing a significand of at most s is $\log_{10} s$.*

2. *Instead of working in base 10, we may work in base B, in which case the Benford probabilities become $\log_B \left(\frac{d+1}{d}\right)$ for the distribution of the first digit, and $\log_B s$ for a significand of at most s.*

Incorporating these two generalizations, we are led to our final definition of Benford's Law.

Definition 1.6.1 (**Strong Benford's Law Base** B)**.** *A data set satisfies the Strong Benford's Law Base* B *if the probability of observing a significand of at most* s *in base* B *is* $\log_B s$*. We shall often refer to the distribution of just the first digit as Benford's Law, as well as the distribution of the entire significand.*

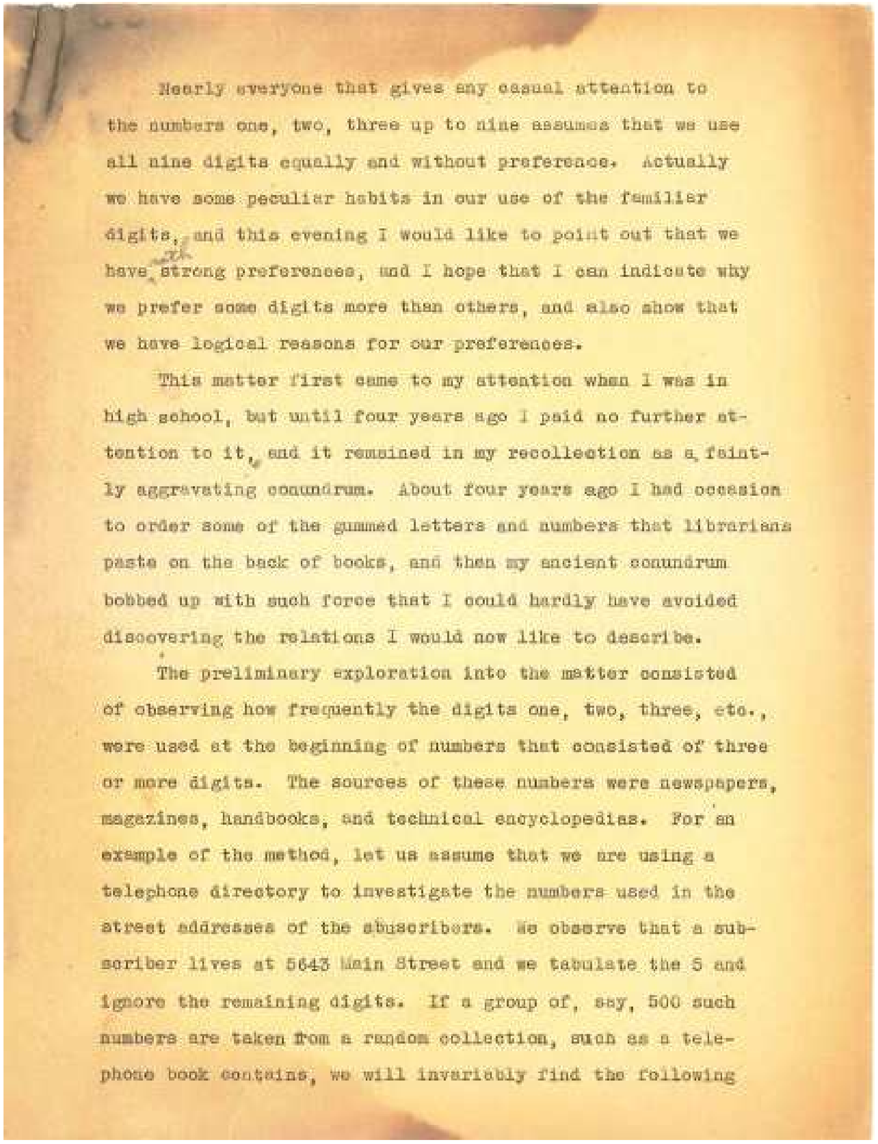
We end the introduction by briefly summarizing the goals of this book and what follows. We address two central questions:

1. *Which data sets (mathematical expressions, physical data, financial transactions) follow this law, and why?*
2. *What are the practical implications of this law?*

There are several different arguments for the first question, depending on the structure of the data. Our studies will show that the answer is deeply connected to results in subjects ranging from probability to Fourier analysis to dynamical systems to number theory. We shall develop enough of these topics for our investigations, recalling standard results in each when needed.

The second question leads to many surprising characters entering the scene. The reason Benford's Law is not just a curiosity of pure mathematics is due to the wealth of applications, in particular to data integrity and fraud tests. There have (sadly) been numerous examples of researchers and corporations tinkering with data; if undetected, the consequences could be severe, ranging from companies not paying their fair share of taxes, to unsafe medical treatments being approved, to unscrupulous researchers being funded at the expense of their honest peers, to electoral fraud and the effective disenfranchisement of voters. With a large enough data set, the laws of probability and statistics state that certain patterns should emerge. Some of these consequences are well known, and thus are easily incorporated by people modifying data. For example, while everyone knows that if you simulate flipping a fair coin 1,000,000 times then there should be about 500,000 heads, fewer know how likely it is to have 100 consecutive heads in the sequence of tosses. The situation is similar with Benford's Law. Almost anyone unfamiliar with Benford's Law would, if asked to simulate data, create a set where either the first digits are equally likely to be anything from 1 to 9, or else clustered around 5. As many real-world data sets follow Benford's Law, this leads to a quick and easy test for fraud. Such tests are now routinely used by the IRS to detect tax fraud, while generalizations may be used in the future to detect whether or not an image has been modified.

What better way to end the introduction than with notes from a talk that Frank Benford gave on the law that now bears his name! While this was one of the earliest talks in the subject, it was by no means the last. As the online bibliography [BerH2] shows, Benford's Law has become a very active research area with numerous applications across disciplines, many of which are described in the following chapters. Enjoy!



Nearly everyone that gives any casual attention to the numbers one, two, three up to nine assumes that we use all nine digits equally and without preference. Actually we have some peculiar habits in our use of the familiar digits, and this evening I would like to point out that we have such strong preferences, and I hope that I can indicate why we prefer some digits more than others, and also show that we have logical reasons for our preferences.

This matter first came to my attention when I was in high school, but until four years ago I paid no further attention to it, and it remained in my recollection as a faintly aggravating conundrum. About four years ago I had occasion to order some of the gummed letters and numbers that librarians paste on the back of books, and then my ancient conundrum bobbed up with such force that I could hardly have avoided discovering the relations I would now like to describe.

The preliminary exploration into the matter consisted of observing how frequently the digits one, two, three, etc., were used at the beginning of numbers that consisted of three or more digits. The sources of these numbers were newspapers, magazines, handbooks, and technical encyclopedias. For an example of the method, let us assume that we are using a telephone directory to investigate the numbers used in the street addresses of the sbuscribers. We observe that a subscriber lives at 5643 Main Street and we tabulate the 5 and ignore the remaining digits. If a group of, say, 500 such numbers are taken from a random collection, such as a telephone book contains, we will invariably find the following

-2-

relations:

The digit one is used 30 per cent of the time as the first digit of numbers and the digit two is used 18 per cent of the time. The percentage falls as we go to higher digits, and the last digit nine is used less than 5 per cent of the time. This is far from an even distribution, and let us yield to a natural curiosity and see if we can discover why the digit one is nearly seven times as popular as nine as a digit with which to begin a number. Since every large group of numbers follows the same rule, we can begin with a generalization about numbers as we use them.

Our natural numbers, one, two, three, etc., are not distributed in the same manner as are natur~~e~~al phenomena and events. In going from one to two, we double the quantity while if we go from nine to ten, we add only eleven per cent to the size of the nine. If, in going from one to two, we take successive increments of eleven per cent of the value of the last number, we find that seven increments or steps are necessary, and this is exactly the occurrence ratio of numbers beginning with the digit one over numbers beginning with the digit nine. Looked at in this way, our so-called natural numbers are poorly adapted for measuring things that can be subdivided into smaller units. Even in our methods of thought, we use a number scale that is a wide departure from the one, two, three, that we are taught is the natural way of counting.

-3-

To illustrate how we actually think when considering sizes, let us assume that you own a row boat that is ten feet long that you think is too small and you therefore want a slightly larger boat. If I offered you an eleven-foot boat, you would possibly be satisfied with this increase. But if you owned a fifty-foot yacht and wished a larger one, I feel quite certain that a fifty-one-foot yacht would not seem significantly longer, but a fifty-five-foot yacht might do. To carry it one step further, an increase of five feet to a five hundred-foot ship would be insignificant, but a fifty-foot increase would be definitely a distinct size of ship. In all three cases the next larger size was a fixed percentage of the previous size, and this is our customary way of thinking of and determining sizes. A logical system of numbers might then run 10, 11 --- 50, 55--- 500, 550 -- -- 1000, 1100 ----

This type of number scale would make arithmetic even less popular in certain quarters than it is now, but people with high school educations are familiar with such a scale. They call it a geometric series, and another name is logarithmic scale.

Our information about the existence of this type of a working scale is based on a tabulation of ~~our~~ over twenty-thousand numbers. It was found that used numbers from every source investigated conformed much more closely to a logarithmic scale than to the natural scale. The final result of this tabulation was that the numbers were found to follow the distribution law

$$F_a = \log \frac{a+1}{a}$$

with a high degree of fidelity.

One of the interesting facts brought out by this investigation was that purely random and unrelated numbers, such as are found in the news items of a newspaper, give better agreement with the formal law than do numbers that are themselves the result of some mathematical law. In the latter class, we can include the tabulations of squares, cubes and square roots of natural numbers given in handbooks. As a result of completely random numbers being more orderly in arrangement than closely related numbers, it has been suggested that this theory be called the Law of Lawless Numbers, and this is essentially the meaning of its formal title, The Law of Anomalous Numbers.

Chapter Two

A Short Introduction to the Mathematical Theory of Benford's Law

Arno Berger and T. P. Hill[1]

[1]Department of Mathematical & Statistical Sciences, University of Alberta, and School of Mathematics, Georgia Institute of Technology.

Chapter Three

Fourier Analysis and Benford's Law

Steven J. Miller[1]

[1]Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267. The author was partially supported by NSF grants DMS0970067 and DMS1265673.

PART II

General Theory II: Distributions and Rates of Convergence

Chapter Four

Benford's Law Geometry

Lawrence Leemis[1]

[1]Department of Mathematics, The College of William & Mary, Williamsburg, VA 23187.

Chapter Five

Explicit Error Bounds via Total Variation

Lutz Dümbgen and Christoph Leuenberger[1]

[1]University of Bern, Bern, Switzerland and University of Fribourg, Fribourg, Switzerland respectively.

Chapter Six

Lévy Processes and Benford's Law

Klaus Schürger[1]

[1]Department of Economics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany.

PART III
Applications I: Accounting and Vote Fraud

Chapter Seven

Benford's Law as a Bridge between Statistics and Accounting

Richard J. Cleary and Jay C. Thibodeau[1]

[1]Mathematics & Science Division, Babson College, and Department of Accountancy, Bentley University.

Chapter Eight

Detecting Fraud and Errors Using Benford's Law

Mark Nigrini[1]

[1]Department of Accounting, West Virginia University, Morgantown, West Virginia 26506.

Chapter Nine

Can Vote Counts' Digits and Benford's Law Diagnose Elections?

Walter R. Mebane, Jr.[1]

[1]Department of Political Science and Department of Statistics, University of Michigan, Ann Arbor, MI. The author thanks Jake Gatof, Joe Klaver, William Macmillan and Matthew Weiss for their assistance.

Chapter Ten

Complementing Benford's Law for Small N: A Local Bootstrap

Boudewijn F. Roukema[1]

[1]Toruń Centre for Astronomy, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, ul. Gagarina 11, 87-100 Toruń, Poland.

PART IV
Applications II: Economics

Chapter Eleven

Measuring the Quality of European Statistics

Bernhard Rauch, Max Göttsche, Gernot Brähler, Stefan Engel[1]

[1]Rauch: University of Regensburg, Department of Economics, Universitätsstraße 31, 93053 Regensburg, Germany; Göttsche and Engel: Catholic University of Eichstätt-Ingolstadt, Department of Auditing and Controlling, Auf der Schanz 49, 85049, Ingolstadt, Germany; Brähler: Ilmenau University of Technology, Department of Taxation Theory and Auditing, Helmholtzplatz 3, 98693 Ilmenau, Germany.

Chapter Twelve

Benford's Law and Fraud in Economic Research

Karl-Heinz Tödter[1]

[1]Research Centre, Deutsche Bundesbank, Frankfurt am Main, Germany

Chapter Thirteen

Testing for Strategic Manipulation of Economic and Financial Data

Charles C. Moul and John V. C. Nye[1]

[1]Economics Department, Farmer School of Business, Miami University, Ohio; and Nye Mercatus Center and Economics Department, George Mason University and National Research University – Higher School of Economics, Moscow, respectively; We thank Marc Taub for excellent research assistance.

PART V
Applications III: Sciences

Chapter Fourteen

Psychology and Benford's Law

Bruce D. Burns and Jonathan Krygier[1]

[1]School of Psychology, The University of Sydney, NSW 2006, Australia. The authors would like to thank Hal Willaby for comments on an earlier draft. The authors were supported by a grant from the University of Sydney, and it is a pleasure to thank them for their generosity.

Chapter Fifteen

Managing Risk in Numbers Games

Mabel C. Chou, Qingxia Kong, Chung-Piaw Teo and Huan Zheng[1]

[1]Chou and Teo: National University of Singapore; Kong: Universidad Adolfo Ibañez; Zheng: Shanghai Jiao Tong University.

Chapter Sixteen

Benford's Law in the Natural Sciences

David Hoyle[1]

[1]Thorpe Informatics Ltd., Adamson House, Towers Business Park, Wilmslow Rd., Manchester, M20 2YY, UK.

Chapter Seventeen

Generalizing Benford's Law

Joanne Lee, Wendy K. Tam Cho, and George Judge[1]

[1]Lee: Researcher, Mathematica Policy Research; Cho: Departments of Political Science and Statistics, and Senior Research Scientist, National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign; Judge: Department of Agricultural and Resource Economics, University of California at Berkeley.

PART VI

Applications IV: Images

Chapter Eighteen

PV Modeling of Medical Imaging Systems

John Chiverton and Kevin Wells[1]

[1]J. Chiverton is with the School of Engineering, University of Portsmouth, UK and K. Wells is with the Centre for Vision, Speech and Signal Processing, University of Surrey, UK. *john.chiverton@port.ac.uk, k.wells@surrey.ac.uk,*

Chapter Nineteen

Application of Benford's Law to Images

Fernando Pérez-González, Tu-Thach Quach, Chaouki T. Abdallah, Gregory L. Heileman and Steven J. Miller[1]

[1]Pérez-González: Department of Signal Theory and Communications, University of Vigo, EE Telecomunicacion, Campus Universitario, 36310 Vigo, Spain; Quach: Sandia National Laboratories, Albuquerque, NM; Miller: Department of Mathematics and Statistics, Williams College, Williamstown, MA; Abdallah and Heileman: Electrical & Computer Engineering Department, University of New Mexico, Albuquerque, NM. Research supported by the European Union under project REWIND (Grant Agreement Number 268478), the European Regional Development Fund (ERDF) and the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), the Galician Regional Government under projects "Consolidation of Research Units" 2009/62, 2010/85 and SCALLOPS (10PXIB322231PR), and the Iberdrola Foundation through the Prince of Asturias Endowed Chair in Information Science and Related Technologies.

PART VII
Exercises

Chapter Twenty

Exercises

20.1 A QUICK INTRODUCTION TO BENFORD'S LAW

A couple of important points.

- There are many problems that would fit in multiple chapters. To help both the instructors and the readers, we have decided to collect them here. Thus, some of the exercises in this chapter will be far more accessible after reading later parts of the book.

- In Mathematica, if you define the following function you can then use it to find the first digit:

  ```
  firstdigit[x_] := Floor[10^Mod[Log[10,x],1]]
  ```

  (a similar function is definable in other languages, but the syntax will differ slightly).

Exercise 20.1.1. *If X is Benford base 10, find the probability that its significand starts 2.789.*

Exercise 20.1.2. *If X is Benford base 10, find the probability that its significand starts with 7.5 (in other words, its significand is in $[7.5, 7.6)$).*

Exercise 20.1.3. *If X is Benford base 10, find the probability that its significand has no 7s in the first k digits (thus a significand of 1.701 would have no 7 in its first digit, but it would have a 7 in its first two digits.*

Exercise 20.1.4. *Consider α^n for various α and various ranges of n; for example, take $\alpha \in \{2, 3, 5, 10, \sqrt{2}, \sqrt{5}, \sqrt{10}, \pi, e, \gamma\}$ (here γ is the Euler–Mascheroni constant; see* `http://en.wikipedia.org/wiki/Euler-Mascheroni_constant` *for a description and properties), and let n go from 1 to N, where $N \in \{10^3, 10^5, 10^7\}$. Which of these data sets do you expect to be Benford? Why or why not? Read up about chi-square goodness of fit tests (see for example* `http://en.wikipedia.org/wiki/Pearson_chi_square`*) and compare the observed frequencies with the Benford probabilities.*

Exercise 20.1.5. *Revisit the previous problem with more values of N. The problem is that there we looked at three snapshots of the behavior; it is far more interesting to plot the chi-square values as a function of N, for N ranging from say 100 to 10^7 or more. You will see especially interesting behavior if you look at the first digits of π^n.*

Exercise 20.1.6. *We have seen that the Benford behavior of a sequence is related to equidistribution of its logarithm. Thus, in the previous problem it may be useful to look at a log-log plot. Thus instead of plotting the chi-square value against the upper bound N, plot the logarithm of the chi-square value against $\log N$.*

Exercise 20.1.7. *Frequently taking logarithms helps illuminate relationships. For example, Kepler's third law (see* `http://www.physicsclassroom.com/class/circles/Lesson-4/Kepler-s-Three-Laws`*) says that the square of the time it takes a planet to orbit a sun is proportional to the cube of the semimajor axis. Find data for these quantities for the eight planets in our system (or nine if you count Pluto!) and plot them, and then do a log-log plot. A huge advantage of log-log plots is that linear relations are easy to observe and estimate; try to find the best fit line here, and note that the slope of the line should be close to 1.5 (if T is the period and L is the length of the semimajor axis, Kepler's third law is that there is a constant C such that $T^2 = CL^3$, or equivalently $T = CL^{3/2}$, or $\log T = \frac{3}{2}\log L + \log C$). Revisit the original plot, and try to see that it supports T^2 is proportional to L^3!*

Exercise 20.1.8. *Prove the log-laws: if $\log_b x_i = y_i$ and $r > 0$ then*

- $\log_b b = 1$ *and* $\log_b 1 = 0$ *(note $\log_b x = y$ means $x = b^y$);*
- $\log_b(x^r) = r\log_b x$*;*
- $\log_b(x_1x_2) = \log_b x_1 + \log_b x_2$ *(the logarithm of a product is the sum of the logarithms);*
- $\log_b(x_1/x_2) = \log_b x_1 - \log_b x_2$ *(the logarithm of a quotient is the difference of the logarithms; this follows directly from the previous two log-laws);*
- $\log_c x = \log_b x/\log_b c$ *(this is the change of base formula).*

Exercise 20.1.9. *The last log-law (the change of base formula) is often forgotten, but is especially important. It tells us that if we can compute logarithms in one base then we can compute them in any base. In other words, it suffices to create just* one *table of logarithms, so we only need to find one base where we can easily compute logarithms. What base do you think that is, and how would you compute logarithms of arbitrary positive real numbers?*

Exercise 20.1.10. *The previous problem is similar to issues that arise in probability textbooks. These books only provide tables of probabilities of random variables drawn from a normal distribution,[1] as one can convert from such a table to probabilities for any other random variable. One such table is online here:*

[1]The random variable X is normally distributed with mean μ and variance σ^2 if its probability density function is $f(x;\mu,\sigma) = \exp\left(-(x-\mu)^2/(2\sigma^2)\right)/\sqrt{2\pi\sigma^2}$.

`http://www.mathsisfun.com/data/standard-normal-distribution-table.html`. *Use a standard table to determine the probability that a normal random variable with mean $\mu = 5$ and variance $\sigma^2 = 16$ (so the standard deviation is $\sigma = 4$) takes on a value between -3 and 7. Thus, similarly to the change of base formula, there is an enormous computational saving as we only need to compute probabilities for one normal distribution.*

Exercise 20.1.11. *Prove* $\frac{d}{dx}\log_b x = \frac{1}{x\log b}$. Hint: *First do this when $b = e$, the base of the natural logarithms; use $e^{\log x} = x$ and the chain rule.*

Exercise 20.1.12. *Revisit the first two problems, but now consider some other sequences, such as $n!$, $\cos(n)$ (in radians of course as otherwise the sequence is periodic), n^2, n^3, $n^{\log n}$, $n^{\log\log n}$, $n^{\log\log\log n}$, n^n.* In some situations $\log_4$ does not mean the logarithm base 4, but rather four iterations of the logarithm function. It might be interesting to investigating $n^{\log_{f(n)} n}$ under this definition for various integer-valued functions f.

Exercise 20.1.13. *Revisit the previous problem but for some recurrence relations. For example, try the Fibonacci numbers ($F_{n+2} = F_{n+1} + F_n$ with $F_0 = 0$ and $F_1 = 1$) and some other relations, such as the following.*

- *Catalan numbers: $C_n = \frac{1}{n+1}\binom{2n}{n}$; these satisfy a more involved recurrence (see* `http://en.wikipedia.org/wiki/Catalan_number`*).*
- *Squaring Fibonaccis: $G_{n+2} = G_{n+1}^2 + G_n^2$ with $G_0 = 0$ and $G_1 = 1$.*
- *F_p where p is a prime (i.e., only look at the Fibonaccis at a prime index).*
- *The logistic map: $x_{n+1} = rx_n(1 - x_n)$ for various choices of r and starting values x_0 (see* `http://en.wikipedia.org/wiki/Recurrence_relation`*).*
- *Newton's method for the* difference *between the nth prediction and the true value. For example, to find the square root of α we use $x_{n+1} = \frac{1}{2}\left(x_n + \frac{\alpha}{x_n}\right)$, and thus we would study the distribution of leading digits of $|\sqrt{\alpha} - x_n|$. One could also look at other roots, other numbers, or more complicated functions. For more on Newton's method, see* `http://mathworld.wolfram.com/NewtonsMethod.html`*.*
- *The $3x + 1$ Map: $x_{n+1} = 3x_n + 1$ if x_n is odd and $x_n/2$ if x_n is even (though some authors use a slightly different definition, where for x_n even, one instead lets $x_{n+1} = x_n/2^d$, where d is the highest power of 2 dividing x_n). It is conjectured that no matter what positive starting seed x_0 you take, eventually x_n cycles among 4, 2, and 1 for n sufficiently large (or is identically 1 from some point onward if we use the second definition). We return to this problem in Chapter 3.*

For the remaining problems, whenever a data set satisfies Benford's Law we mean the *strong* version of the law. This means the cumulative distribution function of the significand is $F_X(s) = \log_{10}(s)$ for $s \in [1, 10)$, which implies that the probability of a first digit of d is $\log_{10}(1 + 1/d)$.

Exercise 20.1.14. *If a data set satisfies (the strong version of) Benford's Law base 10, what are the probabilities of all pairs of leading digits? In other words, what is the probability the first two digits are $d_1 d_2$ (in that order)? What if instead our set were Benford base b?*

Exercise 20.1.15. *Let X be a random variable that satisfies (the strong version of) Benford's Law. What is the probability that the second digit is d? Note here that the possible values of d range from 0 to 9.*

Exercise 20.1.16. *Building on the previous problem, compute the probability that a random variable satisfying the strong version of Benford's Law has its kth digit equal to d. If we denote these probabilities by $p_k(d)$, what is $\lim_{k\to\infty} p_k(d)$? Prove your claim.*

Exercise 20.1.17. *Find a data set that is spread over several orders of magnitude, and investigate its Benfordness (for example, stock prices or volume traded on a company that has been around for decades).*

Exercise 20.1.18. *Look at some of the data sets from the previous exercises that were* not *Benford, and see what happens if you multiply them together. For example, consider $n^2 \cdot \cos(n)$ (in radians), or $n^2\sqrt{10}^n \cos(n)$, or even larger products. Does this support the claim in the chapter that products of random variables tend to converge to Benford behavior?*

Exercise 20.1.19. *Let $\mu_{k;b}$ denote the mean of significands of k digits of random variables perfectly satisfying Benford's Law, and let μ_b denote the mean of the significands of random variables perfectly following Benford's Law. What is $\mu_{k;b}$ for $k \in 1, 2, 3$? Does $\mu_{k;b}$ converge to μ_b? If yes, bound $|\mu_{k;b} - \mu_b|$ as a function of k.*

Exercise 20.1.20. *Benford's Law can be viewed as the distribution on significands arising from the density $p(x) = \frac{1}{x\log(10)}$ on $[1, 10)$ (and 0 otherwise). More generally, consider densities $p_r(x) = C_r/x^r$ for $x \in [1, 10)$ and 0 otherwise with $r \in (-\infty, \infty)$, where C_r is a normalization constant so that the density integrates to 1. For each r, calculate the probability of observing a first digit of d, and calculate the expected value of the first digit.*

20.2 A SHORT INTRODUCTION TO THE MATHEMATICAL THEORY OF BENFORD'S LAW

For a more detailed development of this material, see *An Introduction to Benford's Law* [BerH5] by Berger and Hill.

20.3 FOURIER ANALYSIS AND BENFORD'S LAW

20.3.1 Problems from Introduction to Fourier Analysis

The following exercises are from the chapter "An Introduction to Fourier Analysis," from the book *An Invitation to Modern Number Theory* (Princeton University Press, Steven J. Miller and Ramin Takloo-Bighash). This chapter is available online on the web page for this book (go to the links for Chapter 3).

Exercise 20.3.1. *Prove e^x converges for all $x \in \mathbb{R}$ (even better, for all $x \in \mathbb{C}$). Show the series for e^x also equals*

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n, \tag{20.1}$$

which you may remember from compound interest problems.

Exercise 20.3.2. *Prove, using the series definition, that $e^{x+y} = e^x e^y$ and calculate the derivative of e^x.*

Exercise 20.3.3. *Let f, g, and h be continuous functions on $[0,1]$, and $a, b \in \mathbb{C}$. Prove*

1. *$\langle f, f\rangle \geq 0$, and equals 0 if and only if f is identically zero;*
2. *$\langle f, g\rangle = \overline{\langle g, f\rangle}$;*
3. *$\langle af + bg, h\rangle = a\langle f, h\rangle + b\langle g, h\rangle$.*

Exercise 20.3.4. *Find a vector $\vec{v} = \binom{v_1}{v_2} \in \mathbb{C}^2$ such that $v_1^2 + v_2^2 = 0$, but $\langle \vec{v}, \vec{v}\rangle \neq 0$.*

Exercise 20.3.5. *Prove x^n and x^m are not perpendicular on $[0,1]$. Find a $c \in \mathbb{R}$ such that $x^n - cx^m$ is perpendicular to x^m; c is related to the projection of x^n in the direction of x^m.*

Exercise 20.3.6 (Important)**.** *Show for $m, n \in \mathbb{Z}$ that*

$$\langle e_m(x), e_n(x)\rangle \;=\; \begin{cases} 1 & \text{if } m = n,\\ 0 & \text{otherwise.}\end{cases} \tag{20.2}$$

Exercise 20.3.7. *Let f and g be periodic functions with period a. Prove $\alpha f(x) + \beta g(x)$ is periodic with period a.*

Exercise 20.3.8. *Prove any function can be written as the sum of an even and an odd function.*

Exercise 20.3.9. *Show*

$$\langle f(x) - \widehat{f}(n)e_n(x), e_n(x)\rangle \;=\; 0. \tag{20.3}$$

This agrees with our intuition: after removing the projection in a certain direction, what is left is perpendicular to that direction.

Exercise 20.3.10. *Prove*

1. $\langle f(x) - S_N(x), e_n(x)\rangle = 0$ *if* $|n| \le N$;
2. $|\widehat{f}(n)| \le \int_0^1 |f(x)|dx$;
3. *Bessel's Inequality: if* $\langle f, f\rangle < \infty$ *then* $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 \le \langle f, f\rangle$;
4. *Riemann–Lebesgue Lemma: if* $\langle f, f\rangle < \infty$ *then* $\lim_{|n|\to\infty} \widehat{f}(n) = 0$ *(this holds for more general* f*; it suffices that* $\int_0^1 |f(x)|dx < \infty$*);*
5. *Assume* f *is differentiable* k *times; integrating by parts, show* $|\widehat{f}(n)| \ll \frac{1}{n^k}$ *and the constant depends only on* f *and its first* k *derivatives.*

Exercise 20.3.11. *Let* $h(x) = f(x) + g(x)$*. Does* $\widehat{h}(n) = \widehat{f}(n) + \widehat{g}(n)$*? Let* $k(x) = f(x)g(x)$*. Does* $\widehat{k}(n) = \widehat{f}(n)\widehat{g}(n)$*?*

Exercise 20.3.12. *If* $\langle f, f\rangle, \langle g, g\rangle < \infty$ *then the dot product of* f *and* g *exists:* $\langle f, g\rangle < \infty$ *(see Remark 11.2.4 of [MiT-B]). Do there exist* $f, g : [0,1] \to \mathbb{C}$ *such that* $\int_0^1 |f(x)|dx, \int_0^1 |g(x)|dx < \infty$ *but* $\int_0^1 f(x)\overline{g}(x)dx = \infty$*? Is* $f \in L^2([0,1])$ *a stronger or an equivalent assumption to* $f \in L^1([0,1])$*?*

Exercise 20.3.13. *Define*

$$A_N(x) \;=\; \begin{cases} N & \text{for } |x| \le \frac{1}{N}, \\ 0 & \text{otherwise.} \end{cases} \tag{20.4}$$

Prove A_N *is an approximation to the identity on* $[-\frac{1}{2}, \frac{1}{2}]$*. If* f *is continuously differentiable and periodic with period* 1*, calculate*

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)A_N(x)dx. \tag{20.5}$$

Exercise 20.3.14. *Let* $A(x)$ *be a non-negative function with* $\int_{\mathbb{R}} A(x)dx = 1$*. Prove* $A_N(x) = N \cdot A(Nx)$ *is an approximation to the identity on* $\mathbb{R}$*.*

Exercise 20.3.15 (Important)**.** *Let* $A_N(x)$ *be an approximation to the identity on* $[-\frac{1}{2}, \frac{1}{2}]$*. Let* $f(x)$ *be a continuous function on* $[-\frac{1}{2}, \frac{1}{2}]$*. Prove*

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)A_N(x)dx \;=\; f(0). \tag{20.6}$$

Exercise 20.3.16. *Prove the two formulas above. The geometric series formula will be helpful:*

$$\sum_{n=N}^{M} r^n \;=\; \frac{r^N - r^{M+1}}{1-r}. \tag{20.7}$$

Exercise 20.3.17. *Show that the Dirichlet kernels are* not *an approximation to the identity. How large are* $\int_0^1 |D_N(x)|dx$ *and* $\int_0^1 D_N(x)^2dx$*?*

Exercise 20.3.18. *Prove the Weierstrass Approximation Theorem implies the original version of the Weierstrass Theorem.*

Exercise 20.3.19. *Let $f(x)$ be periodic function with period 1. Show*

$$S_N(x_0) \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)D_N(x-x_0)dx \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x_0-x)D_N(x)dx. \qquad (20.8)$$

Exercise 20.3.20. *Let $\widehat{f}(n)=\frac{1}{2^{|n|}}$. Does $\sum_{-\infty}^{\infty}\widehat{f}(n)e_n(x)$ converge to a continuous, differentiable function? If so, is there a simple expression for that function?*

Exercise 20.3.21. *Fill in the details for the above proof. Prove the result for all f satisfying $\int_0^1 |f(x)|^2dx<\infty$.*

Exercise 20.3.22. *If $\int_0^1 |f(x)|^2dx<\infty$, show Bessel's Inequality implies there exists a B such that $|\widehat{f}(n)|\le B$ for all n.*

Exercise 20.3.23. *Though we used $|a+b|^2\le 4|a|^2+4|b|^2$, any bound of the form $c|a|^2+c|b|^2$ would suffice. What is the smallest c that works for all $a,b\in\mathbb{C}$?*

Exercise 20.3.24. *Let $f(x)=\frac{1}{2}-|x|$ on $[-\frac{1}{2},\frac{1}{2}]$. Calculate $\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}$. Use this to deduce the value of $\sum_{n=1}^{\infty}\frac{1}{n^2}$. This is often denoted $\zeta(2)$ (see Exercise 3.1.7 of [MiT-B]). See [BoPo][2] for connections with continued fractions, and [Kar][3] for connections with quadratic reciprocity.*

Exercise 20.3.25. *Let $f(x)=x$ on $[0,1]$. Evaluate $\sum_{n=1}^{\infty}\frac{1}{n^2}$.*

Exercise 20.3.26. *Let $f(x)=x$ on $[-\frac{1}{2},\frac{1}{2}]$. Prove $\frac{\pi}{4}=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{(2n-1)^2}$. See also Exercise 3.3.29, and see Chapter 11 of [BorB][4] or [Schum][5] for a history of calculations of π.*

Exercise 20.3.27. *Find a function to determine $\sum_{n=1}^{\infty}\frac{1}{n^4}$.*

Exercise 20.3.28. *Show the Gaussian $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-(x-\mu)^2/2\sigma^2}$ is in $\mathcal{S}(\mathbb{R})$ for any $\mu,\sigma\in\mathbb{R}$.*

Exercise 20.3.29. *Let $f(x)$ be a Schwartz function with compact support contained in $[-\sigma,\sigma]$ and denote its Fourier transform by $\widehat{f}(y)$. Prove for any integer $A>0$ that $|\widehat{f}(y)|\le c_f y^{-A}$, where the constant c_f depends only on f, its derivatives and σ. As such a bound is useless at $y=0$, one often derives bounds of the form $|\widehat{f}(y)|\le\frac{\widetilde{c_f}}{(1+|y|)^A}$.*

---

[2] E. Bombieri and A. van der Poorten, *Continued fractions of algebraic numbers*. Pages 137–152 in *Computational Algebra and Number Theory (Sydney, 1992)*, Mathematical Applications, Vol. 325, Kluwer Academic, Dordrecht, 1995.

[3] A. Karlsson, *Applications of heat kernels on Abelian groups: $\zeta(2n)$, quadratic reciprocity, Bessel integral*, preprint.

[4] J. Borwein and P. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, John Wiley and Sons, New York, 1987.

[5] P. Schumer, *Mathematical Journeys*, Wiley-Interscience, John Wiley & Sons, New York, 2004.

Exercise 20.3.30. *Consider*

$$f(x) \;=\; \begin{cases} n^6\left(\frac{1}{n^4} - |n-x|\right) & \text{if } |x-n| \le \frac{1}{n^4} \text{ for some } n \in \mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases} \tag{20.9}$$

Show $f(x)$ is continuous but $F(0)$ is undefined. Show $F(x)$ converges and is well defined for any $x \notin \mathbb{Z}$.

Exercise 20.3.31. *If $g(x)$ decays like $x^{-(1+\eta)}$ for some $\eta > 0$, then $G(x) = \sum_{n\in\mathbb{Z}} g(x+n)$ converges for all x, and is continuous.*

Exercise 20.3.32. *For what weaker assumptions on f, f', f'' does $\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)$?*

Exercise 20.3.33. *One cannot always interchange orders of integration. For simplicity, we give a sequence a_{mn} such that $\sum_m(\sum_n a_{m,n}) \neq \sum_n(\sum_m a_{m,n})$. For $m, n \ge 0$ let*

$$a_{m,n} \;=\; \begin{cases} 1 & \text{if } n = m, \\ -1 & \text{if } n = m+1, \\ 0 & \text{otherwise.} \end{cases} \tag{20.10}$$

Show that the two different orders of summation yield different answers (the reason for this is that the sum of the absolute value of the terms diverges).

Exercise 20.3.34. *Find a family of functions $f_n(x)$ such that*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)dx \;\neq\; \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(x)dx \tag{20.11}$$

and each $f_n(x)$ and $f(x)$ is continuous and $|f_n(x)|, |f(x)| \le M$ for some M and all x.

Exercise 20.3.35. *Let f, g be continuous functions on $I = [0,1]$ or $I = \mathbb{R}$. Show if $\langle f, f\rangle, \langle g, g\rangle < \infty$ then $h = f * g$ exists.* Hint: *Use the Cauchy–Schwarz inequality. Show further that $\widehat{h}(n) = \widehat{f}(n)\widehat{g}(n)$ if $I = [0,1]$ or if $I = \mathbb{R}$. Thus the Fourier transform converts convolution to multiplication.*

Exercise 20.3.36. *Let X_1, X_2 be independent random variables with density p. Prove*

$$\mathrm{Prob}(X_1 + X_2 \in [a,b]) \;=\; \int_a^b (p*p)(z)dz. \tag{20.12}$$

Exercise 20.3.37 (Important)**.** *If for all $i = 1, 2, \dots$ we have $\langle f_i, f_i\rangle < \infty$, prove for all i and j that $\langle f_i * f_j, f_i * f_j\rangle < \infty$. What about $f_1 * (f_2 * f_3)$ (and so on)? Prove $f_1 * (f_2 * f_3) = (f_1 * f_2) * f_3$. Therefore convolution is associative, and we may write $f_1 * \cdots * f_N$ for the convolution of N functions.*

Exercise 20.3.38. *Suppose $X_1, \dots, X_N$ are i.i.d.r.v. from a probability distribution p on $\mathbb{R}$. Determine the probability that $X_1 + \cdots + X_N \in [a,b]$. What must be assumed about p for the integrals to converge?*

Exercise 20.3.39. *One useful property of the Fourier transform is that the derivative of $\widehat{g}$ is the Fourier transform of $2\pi ixg(x)$; thus, differentiation (hard) is converted to multiplication (easy). Explicitly, show*

$$\widehat{g}'(y) \;=\; \int_{-\infty}^{\infty} 2\pi ix \cdot g(x) e^{-2\pi ixy} dx. \tag{20.13}$$

If g is a probability density, note $\widehat{g}'(0) = -2\pi i\mathbb{E}[x]$ and $\widehat{g}''(0) = -4\pi^2\mathbb{E}[x^2]$.

Exercise 20.3.40. *If $B(x) = A(cx)$ for some fixed $c \neq 0$, show $\widehat{B}(y) = \frac{1}{c}\widehat{A}\left(\frac{y}{c}\right)$.*

Exercise 20.3.41. *Show that if the probability density of $X_1 + \cdots + X_N = x$ is $(p * \cdots * p)(x)$ (i.e., the distribution of the sum is given by $p * \cdots * p$), then the probability density of $\frac{X_1+\cdots+X_N}{\sqrt{N}} = x$ is $(\sqrt{N}p * \cdots * \sqrt{N}p)(x\sqrt{N})$. By Exercise 20.3.40, show*

$$FT\left[(\sqrt{N}p * \cdots * \sqrt{N}p)(x\sqrt{N})\right](y) \;=\; \left[\widehat{p}\left(\frac{y}{\sqrt{N}}\right)\right]^N. \tag{20.14}$$

Exercise 20.3.42. *Show for any fixed y that*

$$\lim_{N\to\infty}\left[1 - \frac{2\pi^2y^2}{N} + O\left(\frac{y^3}{N^{3/2}}\right)\right]^N \;=\; e^{-2\pi^2y^2}. \tag{20.15}$$

Exercise 20.3.43. *Show that the Fourier transform of $e^{-2\pi^2y^2}$ at x is $\frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$.* Hint: *This problem requires contour integration from complex analysis.*

Exercise 20.3.44. *Modify the proof to deal with the case of p having mean μ and variance σ^2.*

Exercise 20.3.45. *For reasonable assumptions on p, estimate the rate of convergence to the Gaussian.*

Exercise 20.3.46. *Let p_1, p_2 be two probability densities satisfying*

$$\int_{-\infty}^{\infty} xp_i(x)dx \;=\; 0,\quad \int_{-\infty}^{\infty} x^2p_i(x)dx \;=\; 1,\quad \int_{-\infty}^{\infty} |x|^3p_i(x)dx \;<\; \infty. \tag{20.16}$$

Consider $S_N = X_1 + \cdots + X_N$, where for each i, X_1 is equally likely to be drawn randomly from p_1 or *p_2. Show the Central Limit Theorem is still true in this case. What if we instead had a fixed, finite number of such distributions $p_1, \dots, p_k$, and for each i we draw X_i from p_j with probability q_j (of course, $q_1 + \cdots + q_k = 1$)?*

Exercise 20.3.47 (Gibbs Phenomenon)**.** *Define a periodic with period 1 function by*

$$f(x) \;=\; \begin{cases} -1 & \text{if } -\frac{1}{2} \leq x < 0, \\ 1 & \text{if } 0 \leq x < \frac{1}{2}. \end{cases} \tag{20.17}$$

Prove that the Fourier coefficients are

$$\widehat{f}(n) \;=\; \begin{cases} 0 & \text{if } n \text{ is even}, \\ \frac{4}{n\pi i} & \text{if } n \text{ is odd}. \end{cases} \tag{20.18}$$

Show that the Nth partial Fourier series $S_N(x)$ converges pointwise to $f(x)$ wherever f is continuous, but overshoots and undershoots for x near 0. Hint: *Express the series expansion for $S_N(x)$ as a sum of sines. Note $\frac{\sin(2m\pi x)}{2m\pi} = \int_0^x \cos(2m\pi t)dt$. Express this as the real part of a geometric series of complex exponentials, and use the geometric series formula. This will lead to*

$$S_{2N-1}(x) \;=\; 8\int_0^x \Re\left(\frac{1}{2i}\frac{e^{4n\pi it}-1}{\sin(2\pi t)}\right)dt \;=\; 4\int_0^x \frac{\sin(4n\pi t)}{\sin(2\pi t)}dt, \qquad (20.19)$$

which is about 1.179 (or an overshoot of about 18%) when $x = \frac{1}{4n\pi}$. What can you say about the Fejér series $T_N(x)$ for x near 0?

Exercise 20.3.48 (Nowhere Differentiable Function)**.** *Weierstrass constructed a continuous but nowhere differentiable function! We give a modified example and sketch the proof. Consider*

$$f(x) \;=\; \sum_{n=0}^{\infty} a^n \cos(2^n \cdot 2\pi x), \quad \frac{1}{2} < a < 1. \qquad (20.20)$$

Show f is continuous but nowhere differentiable. Hint: *First show $|a| < 1$ implies f is continuous. Our claim on f follows from noting that if a periodic continuous function g is differentiable at x_0 and $\widehat{g}(n) = 0$ unless $n = \pm 2^m$, then there exists C such that for all n, $|\widehat{g}(n)| \le Cn2^{-n}$. To see this, it suffices to consider $x_0 = 0$ and $g(0) = 0$. Our assumptions imply that $(g, e_m) = 0$ if $2^{n-1} < m < 2^{n+1}$ and $m \neq 2^n$. We have $\widehat{g}(2^n) = (g, e_{2^n}F_{2^{n-1}}(x))$ where F_N is the Fejér kernel. The claim follows from bounding the integral $(g, e_{2^n}F_{2^{n-1}}(x))$. In fact, more is true: Baire showed that, in a certain sense, "most" continuous functions are nowhere differentiable! See, for example, [Fol].*[6]

Exercise 20.3.49 (Isoperimetric Inequality)**.** *Let $\gamma(t) = (x(t), y(t))$ be a smooth closed curve in the plane; we may assume it is parametrized by arc length and has length 1. Prove the enclosed area A is largest when $\gamma(t)$ is a circle.* Hint: *By Green's Theorem*

$$\oint_\gamma xdy - ydx \;=\; 2\mathrm{Area}(A). \qquad (20.21)$$

The assumptions on $\gamma(t)$ imply $x(t), y(t)$ are periodic functions with Fourier series expansions and $\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = 1$. Integrate this equality from $t = 0$ to $t = 1$ to obtain a relation among the Fourier coefficients of $\frac{dx}{dt}$ and $\frac{dx}{dt}$ (which are related to those of $x(t)$ and $y(t)$); (20.21) *gives another relation among the Fourier coefficients. These relations imply $4\pi\mathrm{Area}(A) \le 1$ with strict inequality unless the Fourier coefficients vanish for $|n| > 1$. After some algebra, one finds this implies we have a strict inequality unless γ is a circle.*

Exercise 20.3.50 (Applications to Differential Equations)**.** *One reason for the introduction of Fourier series was to solve differential equations. Consider the vibrating string problem: a unit string with endpoints fixed is stretched into some*

[6] G. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd edition, Pure and Applied Mathematics, Wiley-Interscience, New York, 1999.

initial position and then released; describe its motion as time passes. Let $u(x,t)$ denote the vertical displacement from the rest position x units from the left endpoint at time t. For all t we have $u(0,t) = u(1,t) = 0$ as the endpoints are fixed. Ignoring gravity and friction, for small displacements Newton's laws imply

$$\frac{\partial^2 u(x,t)}{\partial x^2} \;=\; c^2 \frac{\partial^2 u(x,t)}{\partial t^2}, \tag{20.22}$$

where c depends on the tension and density of the string. Guessing a solution of the form

$$u(x,t) \;=\; \sum_{n=1}^{\infty} a_n(t) \sin(n\pi x), \tag{20.23}$$

solve for $a_n(t)$.

One can also study problems on $\mathbb{R}$ by using the Fourier transform. Its use stems from the fact that it converts multiplication to differentiation, and vice versa: if $g(x) = f'(x)$ and $h(x) = xf(x)$, prove that $\widehat{g}(y) = 2\pi i y \widehat{f}(y)$ and $\frac{d\widehat{f}(y)}{dy} = -2\pi i \widehat{h}(y)$. This and Fourier inversion allow us to solve problems such as the heat equation

$$\frac{\partial u(x,t)}{\partial t} \;=\; \frac{\partial^2 u(x,t)}{\partial x^2}, \quad x \in \mathbb{R},\ t > 0 \tag{20.24}$$

with initial conditions $u(x,0) = f(x)$.

20.3.2 Problems from Chapter 1: Revisited

Many of the problems from Chapter 1 are appropriate here as well. In addition to reexamining those problems, consider the following.

Exercise 20.3.51. *Is the sequence $a_n = n^{\log n}$ Benford?*

Exercise 20.3.52. *In some situations $\log_4$ does not mean the logarithm base 4, but rather four iterations of the logarithm function. Investigate $n^{\log_{f(n)} n}$ under this definition for various integer-valued functions f.*

20.3.3 Problems from Chapter 3

Exercise 20.3.53. *Assume an infinite sequence of real numbers $\{x_n\}$ has its logarithms modulo 1, $\{y_n = \log_{10} x_n \bmod 1\}$, satisfying the following property: as $n \to \infty$ the proportion of y_n in any interval $[a,b] \subset [0,1]$ converges to $b-a$* **if** *$b - a > 1/2$. Prove or disprove that $\{x_n\}$ is Benford.*

Exercise 20.3.54. *As $\sqrt{2}$ is irrational, the sequence $\{x_n = n\sqrt{2}\}$ is uniformly distributed modulo 1. Is the sequence $\{x_n^2\}$ uniformly distributed modulo 1?*

Exercise 20.3.55. *Does there exist an irrational α such that α is a root of a quadratic polynomial with integer coefficients* and *the sequence $\{\alpha^n\}_{n=1}^{\infty}$ is Benford base 10?*

Exercise 20.3.56. *We showed a geometric Brownian motion is a Benford-good process; is the sum of two independent geometric Brownian motions Benford-good?*

The next few questions are related to a map we now describe. We showed that, suitably viewed, the $3x + 1$ map leads to Benford behavior (or is close to Benford for almost all large starting seeds). Consider the following map. Let $R(x)$ be the number formed by writing the digits of x in reverse order. If $R(x) = x$ we say x is palindromic. If x is not a palindromic number set $P(x) = x + R(x)$, and if x is palindromic let $P(x) = x$. For a given starting seed x_0 consider the sequence where $x_{n+1} = P(x)$. It is not known whether there are any x_0 such that the resulting sequence diverges to infinity, though it is believed that almost all such numbers do. The first candidate to escape is 196; for more see `http://en.wikipedia.org/wiki/Lychrel_number` (this process is also called "reverse-and-add," and the candidates are called Lychrel numbers).

Exercise 20.3.57. *Consider the reverse-and-add map described above applied to a large starting seed. Find as good of a lower bound as you can for the number of seeds between 10^n and 10^{n+1} such that the resulting sequence stabilizes (i.e., we eventually hit a palindrome).*

Exercise 20.3.58. *Come up with a model to estimate the probability a given starting seed in 10^n and 10^{n+1} has its iterates under the reverse-and-add map diverge to infinity.* Hint: *x plus $R(x)$ is a palindrome if and only if there are no carries when we add; thus you must estimate the probability of having no carries.*

Exercise 20.3.59. *Investigate the Benfordness of sequences arising from the reverse-and-add map for various starting seeds. Of course the calculation is complicated by our lack of knowledge about this map, specifically we don't know even one starting seed that diverges! Look at what happens with various Lychrel numbers. For each N can you find a starting seed x_0 such that it iterates to a palindrome after N or more steps?*

Exercise 20.3.60. *Redo the previous three problems in different bases. Your answer will depend now on the base; for example, much more is known base 2 (there we can give specific starting seeds that iterate to infinity).*

Exercise 20.3.61. *Use the Erdös–Turan Inequality to calculate upper bounds for the discrepancy for various sequences, and use those results to prove Benford behavior. Note you need to find a sequence where you can do the resulting computation. For example, earlier we investigated $a_n = n^{\log n}$; are you able to do the summation for this case?*

Exercise 20.3.62. *Consider the analysis of products of random variables. Fix a probability p (maybe $p = 1/2$), and independent identically distributed random variables $X_1, \dots, X_n$. Assume as $n \to \infty$ the product of the X_i's becomes Benford. What if now we let $\widetilde{\mathcal{X}}_n$ be the random variable where we toss n independent coins, each with probability p, and if the ith toss is a head then X_i is in the product (if the product is empty we use the standard convention that it is then 1). Is this process Benford?*

Exercise 20.3.63. *Redo the previous problem, but drop the assumption that the random variables are identically distributed.*

Exercise 20.3.64. *Redo the previous two problems, but now allow the probability that the ith toss is a head to depend on i.*

Exercise 20.3.65. *Consider*

$$\phi_m \;=\; \begin{cases} m & \text{if } |x-\frac{1}{8}| \le \frac{1}{2m}, \\ 0 & \text{otherwise;} \end{cases} \tag{20.25}$$

this is the function from Example 3.3.5 and led to non-Benford behavior for the product. Can you write down the density for the product?

Exercise 20.3.66. *In the spirit of the previous problem, find other random variables where the product is not Benford.*

Exercise 20.3.67. *Consider a Weibull random variable with a scale parameter α of 1 and translation parameter β of 0; so $f(x;\gamma) = x^{\gamma-1}\exp(x^{\gamma})$ for $x \ge 0$ and is zero otherwise. Investigate the Benfordness of chaining random variables here, where the shape parameter γ is the output of the previous step.*

Exercise 20.3.68. *The methods of [JaKKKM] led to good bounds for chaining exponential and uniform random variables. Can you obtain good, explicit bounds in other cases? For example, consider a binomial process with fixed parameter p.*

Exercise 20.3.69. *Apply the methods of Cuff, Lewis, and Miller (for the Weibull distribution) to other random variables. Consider the generalized Gamma distribution (see*

`http://en.wikipedia.org/wiki/Generalized_gamma_distribution`

for more information), where the density is

$$f(x;a,d,p) \;=\; \frac{p/d^a}{\Gamma(d/p)} x^{d-1} \exp\left(-(x/a)^p\right)$$

for $x > 0$ and 0 otherwise, where a, d, p are positive parameters.

For the next few problems, let $f_r(x) - 1/(1+|x|^r)$ with $r > 1$.

Exercise 20.3.70. *Show that for $r > 1$, $\int_{-\infty}^{\infty} f_r\,(x)\,dx$ is finite, and $\int_{-\infty}^{\infty} f_r\,(x)\,dx$ $= \frac{2\pi}{r}\csc\left(\frac{\pi}{r}\right)$.*

Exercise 20.3.71. *Verify the Fourier transform identity used in our analysis:*

$$p_r\left(e^{b+y}\right)e^{b+y} \;=\; \frac{1}{2}\sin\left(\frac{\pi}{r}\right)e^{2\pi iby}\csc\left(\frac{\pi}{r}\left(1-2\pi iy\right)\right),$$

where $b \in [0,1]$.

20.4 BENFORD'S LAW GEOMETRY

Exercise 20.4.1. *Perform a chi-square goodness-of-fit test on the data values in Table 4.1.*

Exercise 20.4.2. *Let the random variable X have the Benford distribution as defined in this chapter. Find $\mathbb{E}[X]$. Next, generate one million Benford random variates and compute their sample mean. Perform this Monte Carlo experiment several times to ensure that the sample means are near $\mathbb{E}[X]$.*

Exercise 20.4.3. *Let $T \sim$ exponential(1). Find the probability mass function of the leading digit to three-digit accuracy. Compare your results to those in Table 4.2.*

Exercise 20.4.4. *Redo the previous exercise, but instead of finding the probability mass function of the leading digit, find the cumulative distribution function of the significand (i.e., find the probability of observing a significand of at most s).*

Exercise 20.4.5. *Determine the set of conditions on a, b, and c associated with $W \sim$ triangular(a, b, c) which result in $T = 10^W$ following Benford's Law.*

Exercise 20.4.6. *Use R to confirm that the cumulative distribution function $F_x(x) = \mathrm{Prob}(X \le x) = \log_{10}(x+1)$ results in a probability mass function that gives the distribution specified in Benford's Law. What is the range of x?*

Exercise 20.4.7. *Use R to determine whether the cumulative distribution function $F_x(x) = \mathrm{Prob}(X \le x) = x^2$ (for some range for x) results in a probability mass function that gives the distribution specified in Benford's Law. If yes, what is the range for x?*

Exercise 20.4.8. *Which of the following distributions of W follow Benford's Law?*

- $f_W(w) \sim U(0, 3.5)$.
- $f_W(w) \sim U(17, 117)$.
- $f_W(w) = w^3 - w^2 + w$ *for* $0 \le w \le 1$*, and* $1 - w^3 + w^2 - w$ *for* $1 \le w \le 2$.
- $f_W(w) = \sqrt{w}$ *for* $0 \le w \le 1$*, and* $1 - \sqrt{w-1}$ *for* $1 \le w \le 2$.

Exercise 20.4.9. *Let b_1 and b_2 be two different integers exceeding 1. Is there a probability density p on an interval I such that if a random variable X has p for its probability density function then X is Benford in both base b_1 and b_2? What if the two bases are allowed to be real numbers exceeding 1? Prove your claims.*

20.5 EXPLICIT ERROR BOUNDS VIA TOTAL VARIATION

Exercise 20.5.1. *Find* $\mathrm{TV}(\sin(x), [-\pi, \pi])$.

Exercise 20.5.2. *Confirm that* $\mathrm{TV}(h, \mathbb{J}) = \mathrm{TV}^+(h, \mathbb{J}) + \mathrm{TV}^-(h, \mathbb{J})$.

Exercise 20.5.3. *Let Y_o and Z be independent random variables such that Y_o has a density f_o with $\mathrm{TV}(f_o) < \infty$ and Z has distribution π. Verify that $Y := Y_o + Z$ has density $f(y) = \int f_o(y - z)\,\pi(dz)$ with $\mathrm{TV}(f) \le \mathrm{TV}(f_o)$.*

Exercise 20.5.4. *Show that an absolutely continuous probability density f on $\mathbb{R}$ satisfies*

$$\mathrm{TV}(f)^2 \ \le \ \int \frac{f'(x)^2}{f(x)}\,dx.$$

Exercise 20.5.5. *Let $\gamma_{a,\sigma}$ be the density of the Gamma distribution $\mathrm{Gamma}(a,\sigma)$ with shape parameter $a > 0$ and scale parameter $\sigma > 0$, i.e.,*

$$\gamma_{a,\sigma}(x) \ = \ \sigma^{-a} x^{a-1} \exp(-x/\sigma)/\Gamma(a)$$

for $x > 0$, and $\gamma_{a,\sigma} = 0$ on $(-\infty, 0]$.

1. *Show that for $a \ge 1$,*

   $$\mathrm{TV}(\gamma_{a,\sigma}) \ = \ \sigma^{-1}\,\mathrm{TV}(\gamma_{a,1}) \quad \textit{and} \quad \mathrm{TV}(\gamma_{a,1}) \ = \ 2((a-1)/e)^{a-1}/\Gamma(a).$$

2. *It is well known that $\Gamma(t+1) = (t/e)^t\sqrt{2\pi t}(1+o(1))$ as $t \to \infty$ (this is Stirling's formula). What does this imply for $\mathrm{TV}(\gamma_{a,1})$? Show that $\mathrm{TV}(\gamma_{a,\sigma}) \to 0$ as $\sqrt{a}\,\sigma \to \infty$ and $a \ge 1$.*

Exercise 20.5.6. *Let X be a strictly positive random variable with density h on $(0,\infty)$. Verify that $Y := \log_B(X)$ has density f given by $f(y) = \log(B)B^y h(B^y)$ for $y \in \mathbb{R}$.*

Exercise 20.5.7. *Let X be a random variable with distribution $\mathrm{Gamma}(a,\sigma)$ for some $a, \sigma > 0$; see Exercise 20.5.5.*

1. *Determine the density $f_{a,\sigma}$ of $Y := \log_B(X)$. Here you should realize that $f_{a,\sigma}(y) = f_{a,1}(y - \log_B(\sigma))$. Show then that*

   $$\mathrm{TV}(f_{a,\sigma}) \ = \ 2\log(B)(a/e)^a/\Gamma(a).$$

   What happens as $a \to \infty$?

2. *To understand why the leading digits of X are far from Benford's Law for large a, verify that $X = \sigma(a + \sqrt{a}Z_a)$ for a random variable Z_a with mean zero and variance one. (Indeed, the density of Z_a converges uniformly to the standard Gaussian density as $a \to \infty$.) Now investigate the distribution of $Y = \log_B(X)$ as $a \to \infty$.*

20.6 LÉVY PROCESSES AND BENFORD'S LAW

Exercise 20.6.1. *Provide an example of a non-continuous cadlag function.*

Exercise 20.6.2. *Prove that a Weiner process is also a Lévy process.*

Exercise 20.6.3. *Prove that a Poisson process is also a Lévy process.*

Exercise 20.6.4. *Prove that the exponential Lévy process* $\{\exp(X_t)\}$ $(t \in \mathbb{R})$ *is a martingale with respect to* $(\mathcal{F}_t) := \sigma\{X_s : s \leq t\}$ *if and only if* $\mathbb{E}[\exp(X_t)] = 1$.

Exercise 20.6.5. *Let* $f(t) = \mathbb{E}[\exp(it\xi)], g(t) = \mathbb{E}[\exp(it\eta)]$ $(t \in \mathbb{R})$ *be the characteristic functions of (real-) valued random variables* ξ, η $(i = \sqrt{-1})$. *Recall that* $\exp(it) = \cos t + i \sin t$ $(t \in \mathbb{R})$ *and* $\mathbb{E}[\exp(it\xi)] := \mathbb{E}[\cos(t\xi)] + i\mathbb{E}[\sin(t\xi)]$ $(t \in \mathbb{R})$. *Finally,* $\overline{a+ib} := a - ib$ $(a, b \in \mathbb{R})$ *denotes the complex conjugate of* $a + ib$. *Note that* $|f|^2(t) = f(t) \cdot \bar{f}(t)$. *Show the following.*

1. f *is continuous,* $f(0) = 1$, *and* $|f(t)| \leq 1,\ t \in \mathbb{R}$.
2. $\bar{f}$ *is a characteristic function.*
3. $f \cdot g$ *is a characteristic function. Hence,* $|f|^2$*is a characteristic function.*
4. *Let* $h_1, h_2, \ldots$ *be characteristic functions. If* $a_1 \geq 0, a_2 \geq 0, \ldots$ *are real numbers such that* $a_1+a_2+\cdots = 1$, *then* $a_1h_1+a_2h_2+\cdots$ *is a characteristic function.*
5. *Show that every characteristic function* h *is non-negative definite, i.e., for all* $n \geq 2$, *real* $t_1, \ldots, t_n$, *and complex* $a_1, \ldots, a_n$ *we have that*
$$\sum_{j=1}^{n}\sum_{k=1}^{n} h(t_j - t_k)a_j\bar{a}_k \geq 0.$$

Exercise 20.6.6. *Show that, for each real number* $p > 0, f(z) := \cos(2\pi pz)$ $(z \in \mathbb{R})$ *is a characteristic function. Deduce that* $g(z) := (\cos(2\pi pz))^2$ $(z \in \mathbb{R})$ *is a characteristic function.*

Exercise 20.6.7. (This exercise gives an example of a characteristic function which "wildly fluctuates.") *It follows from Exercises 20.6.6 and 20.6.5(4) that*

$$h(z) \ := \ \sum_{k=1}^{\infty} 2^{-k}(\cos(2\pi 7^k z))^2, \quad z \in \mathbb{R}$$

is a characteristic function. Show that h *is of infinite total variation over each non-degenerate interval* $[a, b]$, *i.e.,*

$$\sup\left\{\sum_{k=1}^{n} |h(z_{k+1}) - h(z_k)|\right\} \ = \ \infty,$$

the supremum taken over all $n \geq 1$ *and real numbers* $a \leq z_1 < z_2 < \cdots < z_{n+1} \leq b$.

Hint: *It suffices to prove the claim for intervals* $[r + 7^{-N}, r + 2 \cdot 7^{-N}]$ *(being convenient for calculations!) where* $N \geq 1$ *is an integer and* $r \geq 0$ *a real number. Let* $k \geq N+1$ *and denote by* $I(k)$ *the set of integers* j *such that* $1+(r+7^{-N})7^k < j \leq ((r+2\cdot7^{-N})7^k)$. *For* $j \in I(k)$ *put* $t_{2j-1}(k) = (j-1/4)7^{-k}, t_{2j}(k) = j\cdot7^{-k}$. *Show, by using the inequalities* $|a + b| \geq |a| - |b|$ *and* $|(\cos b)^2 - (\cos a)^2| \leq 2|b - a|$ $(a, b \in \mathbb{R})$ *that*

$$\sum_{j\in I(k)} |h(t_{2j}(k)) - h(t_{2j-1}(k))| \ \geq \ 2(1 - \pi/5)7^{-N}(7/2)^k + \text{const}.$$

Exercise 20.6.8. *1. Try to guess how the integral $\int_a^b f(z)\exp(itz)dz$ behaves as $t \to \infty$ if $f : [a,b] \to \mathbb{R}$ is a step function of the form $f(t) = \sum_{j=1}^m c_j \mathbb{I}_{[b_{j-1},b_j)}(t)$ where $a \le b_0 < b_1 < \cdots < b_m \le b$.*

2. *Verify your guess when f is an indicator function of an interval.*

3. *How does the above integral behave when f is continuous on $[a,b]$?*

Exercise 20.6.9. *Show that a Lévy measure Q satisfies $Q(\mathbb{R} \setminus (-\alpha,\alpha)) < \infty$ for all $\alpha > 0$.*

Exercise 20.6.10. *Let X be a Lévy process having Lévy measure Q. Show that, for fixed $c > 0$ and $s \ge 0$, the process X^* given by $X_t^* = X_{ct+s} - X_s$ $(t \ge 0)$ is a Lévy process having Lévy measure $Q^* = cQ$.*

Exercise 20.6.11. *Let $N = (N_t)$ $(t \ge 0)$ be a Poisson process with parameter $\lambda > 0$.*

1. *Verify that the generating triple of N is given by $(\lambda, 0, Q^*)$ where Q^* has total mass λ concentrated on $\{1\}$.*

2. *Verify (6.15) directly for $X = N$, i.e.,*

$$Q^*(A) \;=\; c^{-1}\mathbb{E}[\#\{s < t \le s+c : \Delta N_t \in A \setminus \{0\}\}]$$

holds for all $c > 0, s \ge 0$, and every Borel set $A \subset \mathbb{R}$.

Exercise 20.6.12. *Let $T_t = \sum_{j=1}^{N_t} \zeta_j$ $(t \ge 0)$ denote the compound Poisson process of Example 6.1.21. (Here, (N_t) is a Poisson process with parameter $\lambda > 0; \zeta_1, \zeta_2, \ldots$ are independent random variables with a common distribution Q_1 such that $Q_1(\{0\}) = 0$. Furthermore, the processes (ζ_n) and (N_t) are independent of each other.)*

1. *Show that the characteristic function g_t of T_t $(t \ge 0)$ is given by*

$$g_t(z) \;=\; \exp\left[\lambda t \int_{\mathbb{R}} (e^{izx} - 1) Q_1(dx)\right]$$

for all $z \in \mathbb{R}$ and $t \ge 0$.

2. *It can be shown (see the reference in Example 6.1.21) that (T_t) is a Lévy process. Determine its generating triple (β, σ^2, Q).*

Exercise 20.6.13. *Let W be a (standard) Brownian motion (BM). Show that, for each $c > 0, W^* = (cW_{t/c^2})$ is a BM (scaling property).*

Exercise 20.6.14. *Let $\xi \sim N(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma > 0$.*

1. *Deduce from (6.26) that the characteristic function of ξ is given by*

$$\mathbb{E}[\exp(iz\xi)] \;=\; \exp(i\mu z - \sigma^2 z^2/2),\; z \in \mathbb{R}.$$

2. *Deduce from the result in (1) that, for all* $\mu, z \in \mathbb{R}$ *and* $\sigma > 0$,

$$\int_{-\infty}^{\infty} \cos(zx)\exp(-(x-\mu)^2/(2\sigma^2))dx \;=\; \sqrt{2\pi\sigma^2}\cos(\mu z)\exp(-\sigma^2 z^2/2)$$

and

$$\int_{-\infty}^{\infty} \sin(zx)\exp(-(x-\mu)^2/(2\sigma^2))dx \;=\; \sqrt{2\pi\sigma^2}\sin(\mu z)\exp(-\sigma^2 z^2/2).$$

Exercise 20.6.15. *Let* $W = (W_t)$ *be a BM. Put*

$$S_{t,u} \;:=\; \sup_{0\le s\le u} |W_{t+s} - W_t|,\; t \ge 0, u > 0.$$

1. *Show that* $S_{t,u}$ *is a random variable. (This requires a little argument since the definition of* $S_{t,u}$ *involves uncountably many random variables!)*
   Hint: *Recall that all sample paths of* W *are continuous.*

2. *Show that* $W_n/n \to 0$ $(n \to \infty)$ *a.s.*

3. *Since, for each fixed* $t \ge 0, (W_{u+t} - W_t)$ $(u \ge 0)$ *is a BM, it follows that*

$$\text{for each } t > 0, S_{t,1} \text{ has the same distribution as } S_{0,1}. \qquad (*)$$

   Furthermore, we have that

$$P(S_{0,1} \ge a) \;\le\; 2\exp(-a^2/2),\; a \ge 0 \qquad (**)$$

   (see, e.g., [KaSh]). Use (2) as well as $(*)$ *and* $(**)$ *to show that*

$$W_t/t \to 0 \;(t \to \infty) \text{ a.s.}$$

   Hint: *Use the Borel–Cantelli Lemma.*

Exercise 20.6.16. *Let* $\xi_1, \xi_2, \ldots$ *be independent random variables defined on some probability space* $(\Omega, \mathcal{F}, P)$, *which have a common distribution given by* $P(\xi_n = +1) = p, P(\xi_n = -1) = 1 - p =: q$ $(n \ge 1)$, *where* $0 < p < 1$. *Put* $S_n := \xi_1 + \cdots + \xi_n,\; n \ge 0$ $(S_0 = 0)$, *and let* $(\mathcal{F}_n)$ $(n \ge 0)$ *be the filtration generated by* (ξ_n). *(Note that* $\mathcal{F}_0 = \{\emptyset, \Omega\}$.*)*

1. *Show that* $Y_n := (q/p)^{S_n}$ $(n \ge 0)$ *is an* $(\mathcal{F}_n)$*-martingale.*

2. *Put* $c(\alpha) := \mathbb{E}[\exp(\alpha\xi_1)] = p\exp(\alpha) + q\exp(-\alpha)$ $(\alpha \in \mathbb{R})$. *Show that, for every fixed* $\alpha \in \mathbb{R}$,

$$Z_n \;:=\; \exp(\alpha S_n)/(c(\alpha))^n \;(n \ge 0)$$

   is an $(\mathcal{F}_n)$*-martingale.*

Exercise 20.6.17. *Let* $\xi_1, \xi_2, \ldots$ *be independent random variables defined on the same probability space, which have a common distribution given by* $P(\xi_n = +1) = P(\xi_n = -1) = 1/2$. *Put* $S_0 = 0$ *and* $S_n = \xi_1 + \cdots + \xi_n$ $(n \ge 1)$ *which means that* (S_n) *is a simple symmetric random walk on* $\mathbb{Z}$, *starting at* 0. *Let* $(\mathcal{F}_n)$ *be the filtration generated by* (ξ_n). *Show that following two sequences are* $(\mathcal{F}_n)$*-martingales:*

1. $(S_n^3 - 3nS_n)$.
2. $(S_n^4 - 6nS_n^2 + 3n^2 + 2n)$.

Hint: *Note that* $\mathbb{E}[\xi_n|\mathcal{F}_{n-1}] = \mathbb{E}[\xi_n] = 0$ *a.s. (since* ξ_n *is independent of* $\mathcal{F}_{n-1}$*), and that* $\mathbb{E}[S_{n-1}^2\xi_n|\mathcal{F}_{n-1}] = S_{n-1}^2\mathbb{E}[\xi_n] = 0$ *a.s. (since* S_{n-1} *is* $\mathcal{F}_{n-1}$*-measurable). Note that* $S_n = S_{n-1} + \xi_n$.

Exercise 20.6.18. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $(\mathcal{F}_n)$ $(n \geq 0)$ *be any filtration on* $(\Omega, \mathcal{F})$*. In the sequel let* $Z = (Z_n)$ $(n \geq 0)$ *and* $H = (H_n)$ $(n \geq 1)$ *be sequences of random variables defined on* $(\Omega, \mathcal{F})$ *such that* Z *is adapted and* H *is predictable which means that, for all* $n \geq 1$, H_n *is* $\mathcal{F}_{n-1}$*-measurable. The sequence* $H \bullet Z$ *given by*

$$(H \bullet Z)_n \;:=\; \sum_{j=1}^{n} H_j(Z_j - Z_{j-1}),\; n \geq 0 \;((H \bullet Z)_0 = 0)$$

is called the H*-transform of* Z *or the (discrete) stochastic integral of* H *with respect to* Z*. Now let* Z *be an* $(\mathcal{F}_n)$*-martingale and assume that* $H_j(Z_j - Z_{j-1}) \in L^1,\; j = 1, 2, \ldots$*. Show that* $H \bullet Z$ *is an* $(\mathcal{F}_n)$*-martingale.*
Hint: *Use the iteration property of conditional expectations (see Example 6.1.29).*

Exercise 20.6.19. *Let* $W = (W_t)$ *be a BM and let* $(\mathcal{F}_t)$ *be the filtration generated by* W*. Show that the following processes are* $(\mathcal{F}_t)$*-martingales:*

1. (W_t).
2. $(W_t^2 - t)$.
3. $(W_t^4 - 6tW_t^2 + 3t^2)$.

Hint: *Note that* $W_t - W_s$ *is independent of* $\mathcal{F}_s$ $(0 \leq s \leq t)$.

Exercise 20.6.20. *Let* (N_t) *be a Poisson process with parameter* $\lambda > 0$*, and put* $M_t = N_t - \lambda t$ $(t \geq 0)$*. Let* $(\mathcal{F}_t)$ *be the filtration generated by* (N_t).

1. *Show that* (M_t) *is an* $(\mathcal{F}_t)$*-martingale.*
   Hint: $N_t - N_s$ *is independent of* $\mathcal{F}_s$ $(0 \leq s < t)$.
2. *Show that* $(M_t^2 - \lambda t)$ *is an* $(\mathcal{F}_t)$*-martingale.*
   Hint: *Write* $M_t^2 - M_s^2 = (M_t - M_s)^2 + 2M_s(M_t - M_s)$ $(0 \leq s < t)$.

Exercise 20.6.21. *Let* (N_t) *be a Poisson process with parameter* $\lambda > 0$*, and let* $c > 0$ *be any constant.*

1. *Determine the constant* $\mu(c)$ *such that the process* $(\exp(cN_t + \mu(c)t))$ $(t \geq 0)$ *is a martingale with respect to the filtration* $(\mathcal{F}_t)$ *generated by* (N_t).
   Hint: *Use Theorem 6.1.30 and Exercise 20.6.11.*
2. *Verify directly that the process obtained in (1) is an* $(\mathcal{F}_t)$*-martingale.*
   Hint: *Use that* $\mathbb{E}[\exp(c(N_t - N_s))|\mathcal{F}_s] = \mathbb{E}[\exp(c(N_t - N_s))]$ *a.s.* $(0 \leq s < t)$ *since* $N_t - N_s$ *is independent of* $\mathcal{F}_s$.

Exercise 20.6.22. *Let ξ have a binomial distribution with parameters $n \geq 1$ and $0 \leq p \leq 1$, i.e.,*

$$P(\xi = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

1. *Use Azuma's inequality (Theorem 6.3.1) to prove the following inequality which is due to H. Chernoff (Ann. Math. Statist.* **23** (1952), 493–507*):*

$$P(|\xi - np| \geq t) \leq 2\exp(-2t^2/n),\ t \geq 0,\ n \geq 1. \qquad (*)$$

   Hint: *ξ has the same distribution as a sum of suitable 0–1 random variables $\xi_1, \ldots, \xi_n$.*

2. *Verify $(*)$ directly for $n = 1$.*

Exercise 20.6.23. *Prove (6.147).*
Hint: *First note that $|g(z)| =: \exp(I(z))$, where*

$$I(z) \;:=\; \int_0^z \frac{\cos x - 1}{x} \left(\log\left(\frac{z}{x}\right)\right)^r dx,\ z \geq 0,\ r > 0.$$

Then (6.147) says that

$$I(z) \leq \frac{1}{2(r+1)} \left(1 - (\log(2z/(3\pi)))^{r+1}\right),\ z \geq 4\pi,\ r > 0. \qquad (*)$$

In order to prove $(*)$ note that the cosine is ≤ 0 on the intervals $J(k) := [(2k-1)\pi - \pi/2, (2k-1)\pi + \pi/2]$, and that

$$J(k) \subset [0, z] \quad \textit{iff} \quad 1 \leq k \leq k(z) \;:=\; \lfloor z/(2\pi) + 1/4 \rfloor. \qquad (**)$$

Hence

$$I(z) \;\leq\; -\sum_{k=1}^{k(z)-1} \int_{J(k)} \frac{1}{x} \left(\log\left(\frac{z}{x}\right)\right)^r dx.$$

Using $(**)$ and comparing with a certain Riemann integral finally yields $(*)$.

Exercise 20.6.24. *A process $Z_t = Z_0 \exp(X_t),\ t \geq 0\ (Z_0 > 0)$ is observed at time points $t = 0, 1, 2, \ldots, T$, where (X_t) is a Lévy process of jump-diffusion type as in Example 6.5.2. Let $H_0(2)$ denote the null hypothesis which says that there exist $\alpha \in \mathbb{R}, c \geq 2, \lambda \geq 0$ and a distribution Q_1 on $\mathbb{R}$ satisfying $Q_1(\{0\}) = 0$ such that (X_t) is associated with α, c, λ, and Q_1. (Note that $H_0(2)$ has a meaning different from that at the beginning of Section 6.5!) Let $H_0(2)$ be rejected if $|\widetilde{L}_T/T - p_{10}(1)| \geq 0.1$ (see (6.100) and (6.150)). Let the level of significance be 0.1. (Note that the rejection of $H_0(2)$ entails the rejection of the null hypothesis that (Z_t) is a Black–Scholes process having volatility ≥ 2; see (6.27).) How large has T to be? (Answer: $T \geq 1715$.)*

Exercise 20.6.25. *A process $Z_t = Z_0 \exp(X_t), t \geq 0\ (Z_0 > 0)$ is observed at the time points $t = 0, 1, 2, \ldots, T$, where $(X_t) = \alpha t + T_t, t \geq 0$. Here, $\alpha \in \mathbb{R}$; (T_t) is a compound Poisson (or CP-)process associated with $\lambda > 0$ and $Q_1 = N(\mu, \sigma^2)$ (see*

Example 6.1.21). Suppose that the null hypothesis $H_0(\lambda^*, \sigma^*)$ $(\lambda^* > 0, \sigma^* > 0)$ *is to be tested, which says that there exist* $\alpha \in \mathbb{R}, \mu \in \mathbb{R}, \lambda \geq \lambda^*$*, and* $\sigma \geq \sigma^*$ *such that* $X_t = \alpha t + T_t$ $(t \geq 0)$*, and* (T_t) *is a CP-process associated with* λ *and* Q_1*. Verify that the test outlined in Exercise 20.6.24, which rejects* $H_0(\lambda^*, \sigma^*)$ *if* $|\widetilde{L}_T/T - p_{10}(1)| \geq 0.1$*, is not applicable no matter how the level of significance* $0 < p_0 < 1$ *is chosen.*
Hint: *Show that there does not exist any (finite) constant* Σ^* *satisfying (6.153)* $(g$ *being the characteristic function of* X_1*,* (X_t) *being an arbitrary Lévy process satisfying* $H_0(\lambda^*, \sigma^*))$*. Use Exercise 20.6.14(2).*

Exercise 20.6.26. *Suppose we observe a process* $Z_t = Z_0 \exp(\mu t + cX_t), t \geq 0$ $(Z_0 > 0)$ *at time points* $t = 0, 1, \ldots, T$*. Let* (X_t) *be a gamma process with parameters* α *and* Δ*, and consider (as in Example 6.5.5) the null hypothesis* $H_0(c^*, \alpha^*, \Delta^*)$ *where* $B = 10, c* = \alpha^* = 1, \Delta^* = 2, p_0 = v = 0.1, m = 1, d_1 = 1$*, and* $\lambda(10) = (2\pi/\log 10)^2$ *(recall that* $\log$ *is the natural logarithm).*

1. *Show that in this special case we can choose* $\Sigma^* = (\log 10)^2/24$*.*

2. *How large has the time horizon* T *to be? (Answer:* $T \geq 2129$ *(instead of* $T \geq 2582$ *as in Example 6.5.5!).)*

Exercise 20.6.27. *Prove the following elementary result (Lemma 6.6.7): Let* $a_1, a_2, \cdots$ *be real numbers such that* $0 \leq a_n < 1$ $(n \geq 1)$ *and* $\sum_{n=1}^{\infty} a_n < \infty$*. Then*

$$\sum_{n=1}^{\infty} a_n^t \to 0 \ (t \to \infty).$$

Exercise 20.6.28. *Prove the claim in Example 6.1.28.*

Exercise 20.6.29. *Prove the iteration property of conditional expectations (see Example 6.1.29).*

Exercise 20.6.30. *Prove Lemma 6.2.1.*

20.7 BENFORD'S LAW AS A BRIDGE BETWEEN STATISTICS AND ACCOUNTING

An auditor decides to run a Benford's Law test on a data set that consists of 1000 legitimate expense records from a business, plus a number of fraudulent transactions that an employee is making to a front for a business set up in a relative's name. Because the employees of the business have to obtain special approval for expenditures over \$10,000, the fraudulent transactions are all for amounts between \$9000 and \$9999. For the 1000 legitimate expenditures, we have this data:

First Digit	Observed
1	314
2	178
3	111
4	92
5	88
6	59
7	56
8	56
9	46

Exercise 20.7.1. *Using the Benford Law test at*

`http://web.williams.edu/Mathematics/sjmiller/public_html/benford /chapter01/MillerNigrini_ExcelBenfordTester_Ver401.xlsx`

(or any other suitable software), verify that the data conforms reasonably well to Benford's Law.

Exercise 20.7.2. *Use trial and error (or some more clever approach) to determine how many fraudulent transactions with first digit 9 would need to be added to the 1000 legitimate observations above in order for the hypothesis that the data follows Benford's Law to be rejected at a five percent significance level. Does this seem plausible?*

Exercise 20.7.3. *What is the role of sample size in the sensitivity of Benford's Law? Suppose there are 10,000 legitimate observations instead of 1000, but the ratios for legitimate observations remains the same, i.e., the number of observations for each digit is multiplied by 10. Try the problem again. What changes?*

Exercise 20.7.4. *In which of the following situations is an auditor most likely to use Benford's Law?*

- *An analysis of a fast food franchise's inventory of hamburgers.*
- *An audit of a Fortune 500 company's monthly total revenue over the fiscal year.*
- *An analysis of a multibillion dollar technology company's significant assets.*

Exercise 20.7.5. *Give an additional example of a way that including Benford's Law in an introductory-level statistics class will meet the four goals of the GAISE report of 2005.*

Exercise 20.7.6. *Determine whether the following situations are Type I errors, Type II errors, or neither.*

- *An auditor uses Benford's Law to analyze the values of canceled checks by a business in the past fiscal year. The auditor finds that there are significant spikes in the data set, with 23 and 37 appearing as the first two digits more often than expected. After further investigation, it was found that there were valid non-fraudulent explanations for the variations in the first digits.*

- *An auditor finds that a company's reported revenue does not follow Benford's Law. Further investigation is taken, and it is found that a manager has been rounding up her weekly sales to the nearest thousand to earn an incentive based on a weekly sales benchmark. The manager claims that the inflated sales were an accounting error.*

- *An owner of a business has falsely claimed to give his employees bonuses on each paycheck based on their monthly sales in order to lower his income taxes. An auditor examines the data, but is unable to confidently claim that the data does not follow Benford's Law. Rather than waste funds on a costly investigation, the auditor chooses not to investigate the owner.*

Exercise 20.7.7. *What are the negative effects of a Type I error in an audit? A Type II error? In what situations might one be more dangerous than the other?*

Exercise 20.7.8. *What are some of the reasons listed in the chapter that might explain why a data set should not be expected to follow Benford's Law?*

Exercise 20.7.9. *Give an example of a reason other than fraud that explains why a data set that is expected to conform to Benford's Law does not.*

20.8 DETECTING FRAUD AND ERRORS USING BENFORD'S LAW

Exercise 20.8.1. *Do the following data sets meet the requirements described by Nigrini in order to be expected to follow Benford's Law? Explain why or why not.*

- *The 4-digit PIN numbers chosen by clients of a local bank.*

- *The annual salaries of graduates from a public university.*

- *Numeric student ID numbers assigned by a school.*

- *The distances in miles between Washington, DC and the 500 most populated cities in the United States (excluding Washington, DC).*

- *Results to a survey of 1000 students asked to provide a number in between 1 and 1,000,000.*

- *The number of tickets bought for all events held in a particular stadium over the past five years.*

Exercise 20.8.2. *Take a company which has been at the heart of a scandal (for example, Enron) and investigate some of its publicly available data.*

Exercise 20.8.3. *An audit of a small company reveals a large number of transactions starting with a 5. Come up with some explanations other than fraud.* Hint: *There are two cases: it is the same amount to the same source each time, and it isn't.*

20.9 CAN VOTE COUNTS' DIGITS AND BENFORD'S LAW DIAGNOSE ELECTIONS?

Exercise 20.9.1. *If X satisfies Benford's Law, then the mean of its second digit is 4.187. What is the mean of the kth digit?*

Exercise 20.9.2. *If X satisfies Benford's Law, multiply by an appropriate power of 10 so that it has k integer digits. What is the probability the last digit is d? What is the probability the last two digits are equal? What is the probability the last two digits differ by 1?*

Exercise 20.9.3. *Find some recent voting data (say city or precinct totals) and investigate the distribution of the first and second digits.*

20.10 COMPLEMENTING BENFORD'S LAW FOR SMALL N: A LOCAL BOOTSTRAP

Exercise 20.10.1. *Do you agree with the assessment that Nigrini's conditions for applying Benford's Law are mostly satisfied? Why or why not?*

Exercise 20.10.2. *Why does having a large $\sigma(\log_{10} x_i)$ and a large $\sigma(\log_{10} w_{i,j})$ ensure that the $v_{i,j}$ first-digit distribution approaches Benford's Law?*

Exercise 20.10.3. *What does it mean for bootstrap methods to be considered "conservative"? Identify some of the ways in which bootstrap methods are conservative.*

Exercise 20.10.4. *There are many conservative statistics. Look up the Bonferroni adjustment for multiple comparisons, as well as alternatives to that.*

Exercise 20.10.5. *How would a local bootstrap realization change if the value of Δ were changed?*

Exercise 20.10.6. *Confirm that if $c_{bK7} > 99.924\%$, then $c_{eK7} > 99.99960\%$.*

20.11 MEASURING THE QUALITY OF EUROPEAN STATISTICS

Exercise 20.11.1. *In which of the following two scenarios would χ^2 be larger?*

- *The first-digit frequencies are mostly identical to the expected Benford distribution, but the digit 1 appears 31.1% of the time and the digit 2 appears 16.6% of the time (compared with the expected values of approximately 30.1% and 17.6%, respectively).*
- *The first-digit frequencies are mostly identical to the expected Benford distribution, but the digit 8 appears 6.12% of the time and the digit 2 appears 3.58% of the time (compared with the expected values of approximately 5.12% and 4.58%, respectively).*

Exercise 20.11.2. *What is μ_b, the value of the mean of the Benford distribution of first digits base b?*

Exercise 20.11.3. *What is the value of a^* if μ_e = 3.5?*

Exercise 20.11.4. *Using Figure 11.1, confirm the values of χ^2, χ^2/n, and d^* for the distribution of first digits for Greek social statistics in the year 2004.*

Exercise 20.11.5. *Using Figure 11.1 and the formula for distance measure a^* used by Judge and Schechter, calculate the value of the mean of the data set (μ_e) in the year 2004. Confirm this value by using the formula $\mu_e = \frac{\Sigma_{i=1}^{9} n\mathrm{Prob}(D_1=i)}{n}$.*

The final problem uses data on two fictitious countries, which is available online

```
http://web.williams.edu/Mathematics/sjmiller/public_html
                    /benford/chapter11/
```

(some of the additional readings on that web page may be useful as well).

Exercise 20.11.6. *Calculate the values χ^2, χ^2/n, d^*, and a^* and compare the results for both countries. Which one of these two countries should be examined closer? Are the outcomes consistent?*

20.12 BENFORD'S LAW AND FRAUD IN ECONOMIC RESEARCH

Exercise 20.12.1. *Use (12.1) to find $f(6)$ and $F(6)$ for Benford's Law.*

Exercise 20.12.2. *If X is a Benford variable defined on $[1, 10)$, then what is the probability that the second digit is 5 given that the first digit is also 5?*

Exercise 20.12.3. *Use (12.4) to confirm that when using Benford's Law for Rounded Figures,* $\mathrm{Prob}(D_1 = 8) = 0.054$.

Exercise 20.12.4. *If X is a Benford variable defined on $[1, 10)$, given that the first digit is 8, what is the probability that the second digit is 0 when rounding to two significant digits? What is the probability that the second digit is 2?*

Exercise 20.12.5. *Using Benford's Law for Rounded Figures as the frequencies of first digits for a data set of 300 observed values, calculate Q_1, Q_2, M_1, and M_2 using (12.6) and (12.7).*

Exercise 20.12.6. *Should the Q_1 test or the M_1 test be used for attempting to detect variations in Benford's Law?*

- *What if the data set in question has a mean of 3.44?*
- *Which test should be used for detecting variations in the Generalized Benford Law?*

Exercise 20.12.7. *The Federal Tax Office (FTO) knows that $\Omega = 10\%$ of tax declarations of small and medium enterprises are falsified. The FTO checks the first digits using Benford's Law. Random samples of tax declarations are drawn and the null hypothesis ($\mathrm{H_o}$) "Conformity to Benford's Law" is tested at the $\alpha = 5\%$ level of significance.*

- *Using (12.9), what rejection rate of $\mathrm{H_o}(\theta)$ would you expect if the probability of a Type II error β lies in the interval [0.05, 0.75]?*
- *The FTO obtained the rejection rate $\theta = 0.12$. Use (12.9) to calculate the probability β of a Type II error.*
- *The FTO arranges for an audit at the taxable enterprise if the Benford test rejects $\mathrm{H_o}$ for a certain tax declaration at the $\alpha = 5\%$ level. What is the probability that such an audit will be provoked erroneously? And what is the probability to forbear an audit erroneously?*

Exercise 20.12.8. *A sample of scientific articles is taken, and 17% are found to have regression coefficients with a doubtful distribution of first digits. Use (12.10) to calculate $\hat{\Omega}$.*

20.13 TESTING FOR STRATEGIC MANIPULATION OF ECONOMIC AND FINANCIAL DATA

Exercise 20.13.1. *What are some of the potential reasons given in Section 13.1 for why data sets that are expected to follow Benford's Law fail to do so?*

Exercise 20.13.2. *Did Benford's Law prove financial misreporting during the financial crisis? Justify your assertion.*

Exercise 20.13.3. *What are some of the potential motives that banks have for manipulating VAR data?*

20.14 PSYCHOLOGY AND BENFORD'S LAW

Exercise 20.14.1. *Using (11.1) in Section 11.3, find χ^2 for the elaborated and unelaborated data from Scott, Barnard, and May's study found in Table 14.1.*

Exercise 20.14.2. *What distribution of leading digits would you expect if people were asked to randomly give an integer from 1 to N? How does your answer depend on N? Try an experiment with some of your friends and family.*

20.15 MANAGING RISK IN NUMBERS GAMES: BENFORD'S LAW AND THE SMALL-NUMBER PHENOMENON

Exercise 20.15.1. *What are the risks associated with a high liability limit in a fixed-odds lottery game? What if the limit is too small?*

Exercise 20.15.2. *From the data obtained in Table 15.1, determine the probability that a given number on a ticket for the UK powerball game is a single digit.*

Exercise 20.15.3. *Figure 15.1 shows the proportion of tickets in a Pennsylvania Pick-3 game with a given first digit. Explain why there are several outliers larger than the mean proportion and no outliers smaller than the mean proportion.*

Exercise 20.15.4. *What is the probability that a Type I player chooses the number 345 in a Pick-3 game?*

Exercise 20.15.5. *Let Alice be a Type II player in a Pick-3 game who bets on a number with three significant digits 80% of the time, a number with two significant digits 15% of the time, and a number with one significant digit 5% of the time. What is the probability that Alice bets on the number 345? The number 45? The number 5?*

Exercise 20.15.6. *In the Pennsylvania Pick-3 game, the least square model indicates that 60.42% of the players are Type I players and 39.58% of the players are Type II players. Based on this model, use (15.4) to calculate the expected proportion of the betting volume on a three-digit number with first significant digit 4.*

Exercise 20.15.7. *Let Bob be a Type II player in a Pick-3 game who bets on a number with three significant digits 80% of the time, but also has a tendency to exhibit switching behavior; that is, he will switch later digits with probability 0.9105, and switch the digit to 0 with probability 0.1054. What is the probability that Bob bets on the number 345?*

Exercise 20.15.8. *Use (15.5) to calculate the probability that Bob chooses a three-digit number in between 520 and 529 inclusive.*

Exercise 20.15.9. *Calculate the variance using the equation in Section 15.4.1 under the scenario that all players randomly select a three-digit number.*

20.16 BENFORD'S LAW IN THE NATURAL SCIENCES

Exercise 20.16.1. *Demonstrate that (16.3) holds for $\alpha = 2$.*

Exercise 20.16.2. *Rewrite the log-normal distribution density function (16.5) as the log-normal density function (16.6).*

Exercise 20.16.3. *Show that as σ grows larger, the log-normal density function approaches the power law $p(x) = C_\sigma x^{-1}$, where C_σ is a constant depending on σ.*

Exercise 20.16.4. *Provide examples not mentioned in the chapter of scientific data sets that are not effectively scale invariant.*

Exercise 20.16.5. *Explain the intuition behind why the following distributions are approximately Benford:*

- *The Boltzman–Gibbs distribution (16.8).*
- *The Fermi–Dirac distribution (16.9).*
- *The Bose–Einstein distribution (16.10).*

Exercise 20.16.6. *Obtain a physics textbook (or a CRC handbook, or...) and find a list of physical constants. Perform a chi-square test to determine whether the list of constants follows Benford's Law as expected.*

Exercise 20.16.7. *Sandon found agreement between Benford's Law and population and surface area data for the countries of the world. Find a source that provides the population density of each country. Then determine if population density follows Benford's Law. This can be done using a chi-square test. In general, should the ratio of two Benford random variables be Benford?*

20.17 GENERALIZING BENFORD'S LAW: A REEXAMINATION OF FALSIFIED CLINICAL DATA

Exercise 20.17.1. *Use (17.1) to calculate the average frequency of first digits in Stigler's distribution of first significant digits. Check to see that the distribution matches the values displayed in Table 17.1.*

Exercise 20.17.2. *Verify (17.3), (17.5), and (17.6). Then verify that the sum of the three subsets matches (17.7).*

Exercise 20.17.3. *Calculate the mean of the Stigler FSD distribution and Benford FSD distribution to confirm that they are equivalent to 3.55 and 3.44, respectively.*

Exercise 20.17.4. *For the Estimated Maximum Entropy FSD distribution for data with an FSD mean of 3.44 shown in Table 17.3, find $H(p)$ and ensure that the criteria from (17.13) and (17.14) are reached.*

- *If the Estimated Maximum Entropy FSD distribution is accurate, then the listed probabilities will maximize $H(p)$. First, determine whether replacing $\hat{p_1}$ with 0.231 and $\hat{p_2}$ with 0.2 still allows (17.13) and (17.14) to hold. Now find $H(p)$. Is $H(p)$ larger or smaller than before?*

Exercise 20.17.5. *If the FSD mean is 5, what will be the estimated maximum entropy FSD distribution? What is $\mathrm{Var}(d)$ according to (17.18)?*

Exercise 20.17.6. *Examining the Poehlman data in Table 17.4, calculate the difference for each digit FSD distribution given by Benford's Law.*

Exercise 20.17.7. *The estimated empirical likelihood distributions given an FSD mean will maximize $\sum_{i=1}^{9} p_i$. To test this, ensure that the product of the p_i's from Table 17.5 are greater than the empirical data found in Table 17.4.*

Exercise 20.17.8. *A researcher is trying to decide whether a data set follows Benford's Law or Stigler's Law. What values of the mean of the leading digit suggest Benford over Stigler? What values suggest Stigler over Benford?*

20.18 PARTIAL VOLUME MODELING OF MEDICAL IMAGING SYSTEMS USING THE BENFORD DISTRIBUTION

Exercise 20.18.1. *What is the PV effect? What implications does the PV effect have for medical imaging?*

Exercise 20.18.2. *Prove Corollary 18.3.4.*

Exercise 20.18.3. *What advantages are there to describing the PV effect using matrices as in (18.11)?*

Exercise 20.18.4. *What are the differences between a Rician noise model described by (18.12) and a Gaussian noise model described in (18.13)?*

Exercise 20.18.5. *Use (18.22) to calculate $p(\alpha)$ for $\alpha = 0.50$, where α has two digits of precision.*

Exercise 20.18.6. *How is the contrast to noise ratio (CNR) affected if both the distance between the signal levels of two components and the standard deviation of each class is doubled?*

20.19 APPLICATION OF BENFORD'S LAW TO IMAGES

Exercise 20.19.1. *In (19.9) one of the factors is $\Gamma\left(\frac{-j2\pi n+\log 10}{c\log 10}\right)$, where $j = \sqrt{-1}$. Estimate how rapidly this tends to zero as $|n| \to \infty$ as a function of c (if you wish, choose some values of c to get a feel of the behavior).*

Exercise 20.19.2. *In (19.19) we find that $|a_n(c,\sigma)| \le |a_n(c^+)|$ for all n; investigate how close these can be for various choices of c and σ.*

Exercise 20.19.3. *In Example 19.5.3 we found four zero-mean Gaussians with shaping parameter $c = 1$ with four different standard deviations and $a_1 = 0$. Can you find six zero-mean Gaussians with shaping parameter $c = 1$ and six different standard deviations with $a_1 = 0$? What about eight? More generally, can you find $2m$ such Gaussians for m a positive integer?*

Bibliography

[Abd] H. Abdi, *Bonferroni and Sidak Corrections for Multiple Comparisons*, (2007), in (N. J. Salkind ed.), Encyclopedia of Measurement and Statistics. Thousand Oaks, USA: Sage.

[AbdTh] M. Abdolmohammadi and J. Thibodeau, *Auditing*, in Encyclopedia of Business and Finance, 2nd edition, Macmillan, New York, 2007.

[AbrSteg] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1964.

[AceSbe] E. D. Acebo and M. Sbert, *Benford's Law for Natural and Synthetic Images*, Workshop on Computational Aesthetics in Graphics, Visualization and Imaging, (2007), 169–176.

[Adh] A. K. Adhikari, *Some Results on the Distribution of the Most Significant Digit*, Sankhyā: Indian Journal of Statistics, Series B **31** (1969), 413–420.

[AdhSa] A. K. Adhikari and B. P. Sarkar, *Distribution of Most Significant Digit in Certain Functions Whose Arguments Are Random Variables*, Sankhyā: Indian Journal of Statistics, Series B **30** (1968), 47–58.

[Al] P. C. Allaart, *An Invariant-Sum Characterization of Benford's Law*, **34** (1997), 288–291.

[AnRoSt] T. Anderson, L. Rolen and R. Stoehr, *Benford's Law for Coefficients of Modular Forms and Partition Functions*, Proc. Am. Math. Soc. **139** (2011), no. 5, 1533–1541.

[An] Anonymous, *Weibull Survival Probability Distribution*, last modified May 11, 2009.
`http://www.wepapers.com/Papers/30999/Weibull_Survival_Probability_Distribution`.

[ArJa] L. Arshadi and A. H. Jahangir, *Benford's Law Behavior of Internet Traffic*, Journal of Network and Computer Applications **40** (2014), 194–205.

[AuCa] M. Auffhammer and R. Carson, *Forecasting the Path of China's CO2 Emissions Using Province Level Information*, Journal of Environmental Economics and Management **47** (2008), no. 1, 47–62.

[AuHi] K. Auspurg, T. Hinz, *What Fuels Publication Bias?*, Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) **231**, no. 5 and 6 (2011), 636–660.

[BaBeMi] S. Baek, S. Bernhardsson and P. Minnhagen, *Zipf's Law Unzipped*, New Journal of Physics **13** (2011), 043004, doi:10.1088/1367-2630/13/4/043004

[Bag] K. A. Baggerly, *Empirical Likelihood as a Goodness-of-Fit Measure*, Biometrika, 1998.

[BaBoDS] J. Baik, A. Borodin, P. Deift and T. Suidan, *A Model for the Bus System in Cuernevaca (Mexico)*, Math. Phys. 2005, 1–9. Online at `http://arxiv.org/abs/math/0510414`.

[BaPiYo] P. Baine, J. Pitman and M. Yor, *Probability Laws Related to the Jacobi Theta and Riemann Zeta Functions, and Brownian Excursions*, Bulletin of the AMS **38**, No. 4, June 2001, 435–465.

[Bak] A. Baker, *The Theory of Linear Forms in Logarithms*, Transcendence Theory: Advances and Applications (A. Baker and D. W. Masser, eds), Academic Press, 1977.

[BarFe] J. A. Bargh and M. J. Ferguson, *Beyond Behaviorism: On the Automaticity of Higher Mental Processes*, Psychological Bulletin **126** (2000), 925–945.

[BaHeSl] T. Barrale, R. Hendel and M. Sluys, *Sequences of Initial Digits of Fibonacci Numbers*, Aportaciones Matematicas, Investigacion **20** (2011), 25–42.

[BaLi] P. Baxandall and H. Liebeck, *Vector Calculus*, Clarendon Press, Oxford, 1986.

[Be] R. Beals, *Notes on Fourier Series*, Lecture Notes, Yale University, 1994.

[BebSca] B. Beber and A. Scacco, *The Devil Is in the Digits*, (2009), Washington Post, 20 June 2009.

[BeEsJ] G. Beeli, M. Esslen and L. Jäncke, *Frequency Correlates in Grapheme-Color Synaesthesia*, Psychological Science **18** (2007), no. 9, 788–792.

[Ben] F. Benford, *The Law of Anomalous Numbers*, Proceedings of the American Philosophical Society **78** (1938), 551–572.

[Ber1] A. Berger, *Chaos and Chance*, deGruyter, Berlin, 2001.

[Ber2] A. Berger, *Multi-dimensional Dynamical Systems and Benford's Law*, Discrete Contin. Dyn. Syst. **13** (2005), 219–237.

[Ber3] A. Berger, *Benford's Law in Power-like Dynamical Systems*, Stoch. Dyn. **5** (2005), 587–607.

[Ber4] A. Berger, *Some Dynamical Properties of Benford Sequences*, J. Difference Equ. Appl. **17** (2011), no. 2, 137–159.

[BerBH] A. Berger, L. A. Bunimovich and T. Hill, *One-Dimensional Dynamical Systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197–219.

[BerH1] A. Berger and T. P. Hill, *Newton's Method Obeys Benford's Law*, American Mathematical Monthly **114** (2007), 588–601.

[BerH2] A. Berger and T. P. Hill, *Benford Online Bibliography*, `http://www.benfordonline.net`.

[BerH3] A. Berger and T. P. Hill, *Benford's Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem*, Mathematical Intelligencer **33** (2011), 85–91.

[BerH4] A. Berger and T. P. Hill, *A Basic Theory of Benford's Law*, Probab. Surv. **8** (2011), 1–126.

[BerH5] A. Berger and T. P. Hill, *An Introduction to Benford's Law*, Princeton University Press, Princeton, 2015.

[BerHKR] A. Berger, T. P. Hill, B. Kaynar and A. Ridder, *Finite-State Markov Chains Obey Benford's Law*, to appear in SIAM J. Matrix Analysis.

[Berg] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd edition, Springer Series in Statistics, 1985.

[BerRin] D. Berman and T. Rintoul, *Preliminary Analysis of the Voting Figures in Iran's 2009 Presidential Election* (Ali Ansari ed.), London: Chatham House, St Andrews, 2009.

[Ber] J. Bertoin, *Lévy Processes*, Cambridge University Press, Cambridge, 1996.

[Bh] R. N. Bhattacharya, *Speed of Convergence of the n-Fold Convolution of a Probability Measure on a Compact Group*, Z. Wahrscheinlichkeitstheorie verw. Geb. **25** (1972), 1–10.

[Bi1] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[Bi2] P. Billingsley, *Prime Numbers and Brownian Motion*, Amer. Math. Monthly **80** (1973), 1099–1115.

[Bi3] P. Billingsley, *Probability and Measure*, 3rd edition, John Wiley & Sons, New York, 1995.

[BlS] F. Black and M. Scholes, *The Prices of Options and Corporate Liabilities*, J. Political Economy **81** (1973), 637–659.

[BlGs] A. Blais and T. Gschwend, *Strategic Defection Across Elections, Parties and Voters*, in Citizens, Context, and Choice: How Context Shapes Citizens' Electoral Choices (R. Dalton and C. Anderson, eds), Oxford University Press, New York, 2011, 176–195.

[BlNa] A. Blais and R. Nadeau, *Measuring Strategic Voting: A Two-Step Procedure*, Electoral Studies **15** (1996), no. 1, 39–52.

[BlNaGN] A. Blais, R. Nadeau, E. Gidengil and N. Nevitte, *Measuring Strategic Voting in Multiparty Plurality Elections*, Electoral Studies **20** (2001), no. 2, 343–352.

[BohJe] H. Bohr and B. Jessen, *On the Distribution of the Values of the Riemann Zeta-Function*, Amer. J. Math. **58** (1936), 35–44.

[BolHa] R. Bolton and D. Hand, *Statistical Fraud Detection: A Review*, Statistical Science **17** (2002), no. 3, 235–255.

[BomPo] E. Bombieri and A. van der Poorten, *Continued Fractions of Algebraic Numbers*, in Computational Algebra and Number Theory (Sydney, 1992), Mathematical Applications, **325**, Kluwer Academic, Dordrecht, 1995, 137–152.

[Bon..] S. Bonhoeffer, A. Herz, M. Boerlijst, S. Nee, M. Nowak and R. May, *No Signs of Hidden Language in Noncoding DNA*, Physical Review Letters **76** (1996), no. 11, 1977.

[BorB] J. Borwein and P. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, John Wiley and Sons, New York, 1987.

[BowAz] A. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Clarendon Press, Oxford, 1997.

[BoyDi] W. Boyce and R. DiPrima, *Elementary Differential Equations and Boundary Value Problems*, 7th edition, John Wiley & Sons, New York, 2000.

[Boy] J. Boyle, *An Application of Fourier Series to the Most Significant Digit Problem*, Amer. Math. Monthly **101** (1994), 879–886.

[Br] L. Breiman, *Probability*, Addison–Wesley, Reading, MA, 1968.

[BrDu] J. Brown and R. Duncan, *Modulo One Uniform Distribution of the Sequence of Logarithms of Certain Recursive Sequences*, Fibonacci Quarterly **8** (1970), 482–486.

[Bro1] R. Brown, *Benford's Law and the Screening of Analytical Data: The Case of Pollutant Concentrations in Ambient Air*, Analyst **130** (2005), 1280–1285.

[Bro2] R. Brown, *The Use of Zipf's Law in the Screening of Analytical Data: A Step Beyond Benford*, Analyst **132** (2007), 344–349.

[Buc] B. Buck, A. Merchant and S. Perez, *An Illustration of Benford's First Digit Law Using Alpha Decay Half-Lives*, European Journal of Physics **14** (1993), 59–63.

[Bun1] Der Bundeswahlleiter, *Ergebnisse der Wahlbezirksstatistik: Die Wahlleiter des Bundes und der Länder, Auszugsweise Vervielfältigung und Verbreitung mit Quellenangaben gestattet*, Wahl zum 16, Deutschen Bundestag am 18, September 2005, Wiesbaden: im Auftrag der Herausgebergemeinschaft.

[Bun2] Der Bundeswahlleiter, *Ergebnisse der Wahlbezirksstatistik: Die Wahlleiter des Bundes und der Länder, Auszugsweise Vervielfältigung und Verbreitung mit Quellenangaben gestattet*, Wahl zum 17, Deutschen Bundestag am 27, September 2009, Wiesbaden: im Auftrag der Herausgebergemeinschaft.

[Bun3] Der Bundeswahlleiter, *Federal Elections Act*, accessed March 21, 2012. URL `http://www.bundeswahlleiter.de/en/bundestagswahlen/downloads/rechtsgrundlage%n/bundeswahlgesetz_engl.pdf`.

[Bun4] Der Bundeswahlleiter, *Ergebnisse der Wahlbezirksstatistik: Die Wahlleiter des Bundes und der Länder, Auszugsweise Vervielfältigung und Verbreitung mit Quellenangaben gestattet*, Wahl zum 15, Deutschen Bundestag am 22, September 2002, Wiesbaden: im Auftrag der Herausgebergemeinschaft.

[Burke] J. Burke and E. Kincanon, *Benford's Law and Physical Constants: The Distribution of Initial Digits*, American Journal of Physics **59** (1991), no. 10, 952.

[BurRA] L. E. Burman, W. R. Reed, J. Alm, *A Call for Replication Studies*, Public Finance Review **38**, no. 6 (2010), 787–793.

[Burns] B. D. Burns, *Sensitivity to Statistical Regularities: People (Largely) Follow Benford's Law*, in (N. A. Taatgen & H. van Rijn, eds.), Proceedings of the 31st Annual Conference of the Cognitive Science Society (2000), 2872–2877. Austin, TX: Cognitive Science Society.

[BurnsCo] B. D. Burns and B. Corpus, *Randomness and Inductions from Streaks: "Gambler's Fallacy" versus "Hot Hand."*, Psychonomic Bulletin & Review **11** (2004), 179–184.

[CanSe] F. Cantu and S. Saiegh, *Fraudulent Democracy? An Analysis of Argentina's Infamous Decade Using Supervised Machine Learning*, Political Analysis **19** (2011), no. 4, 409–433.

[Car] L. Carleson, *On the Convergence and Growth of Partial Sums of Fourier Series*, Acta Math. **116** (1966), 135–157.

[Carr] K. J. Carroll, *On the Use and Utility of the Weibull Model in the Analysis of Survival Data*, Controlled Clinical Trials **24** (2003), no. 6, 682–701.

[Cars] C. A. Carslaw, *Anomalies in Income Numbers: Evidence of Goal-Oriented Behavior*, Accounting Review **LXIII** (1988), no. 2, 321–327.

[Cart] Carter Center, *Observing the Venezuela Presidential Recall Referendum: Comprehensive Report*, 2005.

[Cas] J. W. S. Cassels, *An Introduction to Diophantine Approximation*, Cambridge University Press, London 1957.

[Cast] J. Castaneda, *Perpetuating Power: How Mexican Presidents Were Chosen*, New Press, New York, 2000.

[CenMA] Center for Morphometric Analysis at Massachusetts General Hospital, "20 Normal MR Brain Data Sets and Their Manual Segmentations", `http://www.cma.mgh.harvard.edu/ibsr/`, Accessed 2006.

[Che] H. Chernoff, *How to Beat the Massachusetts Number Game: An Application of Some Basic Ideas in Probability and Statistics*, Mathematical Intelligencer **3** (1999), 166–175.

[ChKo] P. Chhibber and K. Kollman, *The Formation of National Party Systems: Federalism and Party Competition in Canada, Great Britain, India, and the United States*, Princeton University Press, Princeton, NJ, 2004.

[ChWe] J. P. Chiverton and K. Wells, *Mixture Effects in FIR Low-Pass Filtered Signals*, IEEE Signal Processing Letters, 13(6), June 2006, 369–372.

[ChGa] W. K. T. Cho and B. J. Gaines, *Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance*, American Statistician **61** (2007), 218–223.

[ChKTWZ] M. C. Chou, Q. X. Kong, C. P. Teo, Z. Z. Wang and H. Zheng, *Benford's Law and Number Selection in Fixed-Odds Numbers Game*, Journal of Gambling Studies **25** (2009), no. 4, 503–521.

[ChT] Y. S. Chow and H. Teicher, *Probability Theory. Independence, Interchangeability, Martingales*, 3rd edition, Springer, 1997.

[ClTh] R. J. Cleary and J. C. Thibodeau, *Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors*, Auditing: A Journal of Practice and Theory **24** (2005), no. 1, 77–81.

[Conr] J. B. Conrey, *The Riemann Hypothesis*, Notices of the AMS, March 2003, 341–353.

[ConT] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall, London, 2004.

[ConG] J. H. Conway and R. Guy, *The Book of Numbers*, Springer, Berlin, 1996.

[CorSo] B. Corominas-Murtra and R. Solé, *Universality of Zipf's Law*, Physical Review E **82** (2010), 011102, doi:10.1103/PhysRevE.82.011102.

[CoEZ] M. Corazza, A. Ellero and A. Zorzi, *Checking Financial Markets via Benford's Law: The S&P 500 Case*. In Mathematical and Statistical Methods for Actuarial Sciences and Finance (Corazza and Pizzi, eds.), Springer, 2010, 93–101.

[CoBr] O. Corzoa and N. Brachob, *Application of Weibull Distribution Model to Describe the Vacuum Pulse Osmotic Dehydration of Sardine Sheets*, LWT - Food Science and Technology **41** (2008), no. 6, 1108–1115.

[CosLTF] E. Costas, V. López-Rodas, F. J. Toro and A. Flores-Moya, *The Number of Cells in Colonies of the Cyanobacterium Microcystis aeruginosa Satisfies Benford's Law*, Aquatic Botany **89** (2008), DOI 10.1016/j.aquabot.2008.03.011.

[CotGr] W. N. Cottingham and D. A. Greenwood, *An Introduction to Nuclear Physics*, Cambridge University Press, Cambridge, 2nd edition, 2001.

[Cox1] G. Cox, *Strategic Voting Equilibria Under the Single Nontransferable Vote*, American Political Science Review **88** (1994), no. 3, 608–621.

[Cox2] G. Cox, *Making Votes Count: Strategic Coordination in the World's Electoral Systems*, Cambridge University Press, New York, 1996.

[CozMil] M. Cozzens and S. J. Miller, *The Mathematics of Encryption: An Elementary Introduction*, AMS Mathematical World series **29**, Providence, RI, 2013, 332 pages.

[CrRe] N. A. C. Cressie and T. R. C. Read, *Multinomial Goodness-of-Fit Tests*, Journal of the Royal Statistics Society Series B, 1984.

[CuLeMi] V. Cuff, A. Lewis and S. J. Miller, *The Weibull Distribution and Benford's Law*, preprint.

[CuKeCh] X. Cui, M. Kerr and G. Churchill, *Transformations for cDNA Microarray Data*, Statistical Applications in Genetics and Molecular Biology **2** (2003), Article 4.

[DM] D. Danilov and J. R. Magnus, *On the Harm That Ignoring Pretesting Can Cause*, Journal of Econometrics **122** (2004), 27–46.

[Da] H. Davenport, *Multiplicative Number Theory*, 2nd edition, Graduate Texts in Mathematics **74**, Springer, New York, 1980, revised by H. Montgomery.

[DaNa] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd edition, Wiley Interscience, Hoboken, NJ, 2003.

[DeDS] M. J. K. De Ceuster, G. Dhaene and T. Schatteman, *On the Hypothesis of Psychological Barriers in Stock Markets and Benford's Law*, Journal of Empirical Finance **5** (1998), no. 3, 263–279.

[DeMyOr] J. Deckert, M. Myagkov and P. Ordeshook, *Benford's Law and the Detection of Election Fraud*, Political Analysis **19** (2011), no. 3, 245–268.

[Deh] S. Dehaene, *The Number Sense: How the Mind Creates Mathematics* (1997), Oxford University Press, Oxford, UK.

[DeHa] S. De Marchi and J. Hamilton, *Trust, but Verify: Assessing the Accuracy of Self-Reported Pollution Data*, Journal of Risk and Uncertainty **32** (2006), no. 1, 57–76.

[Dia] P. Diaconis, *The Distribution of Leading Digits and Uniform Distribution mod 1*, Ann. Probab. **5** (1979), 72–81.

[DiaS-C] P. Diaconis and L. Saloff-Coste, *Comparison Theorems for Reversible Markov Chains*, Ann. Appl. Probab. **3** (1993), no. 3, 696–730.

[Die] A. Diekmann, *Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data*, Journal of Applied Statistics **34** (2007), no. 3, 321–329.

[DieJa] A. Diekmann and B. Jann, *Benford's Law and Fraud Detection: Facts and Legends*, German Economic Review **11** (2010), no. 3, 397–401.

[Dir] P. Dirac, *Cosmological Models and the Large Numbers Hypothesis*, Proceedings of the Royal Society of London A **338** (1974), no. 1615, 439–446.

[DoMeOl] S. Dorogovtsev, J. Mendes and J. Oliveira, *Frequency of Occurrence of Numbers in the World Wide Web*, Physica A **360** (2006), 548–556.

[DrNi] P. D. Drake and M. J. Nigrini, *Computer Assisted Analytical Procedures Using Benford's Law*, Journal of Accounting Education **18** (2000), 127–146.

[DuFr] L. Dubins and D. Freedman, *Random Distribution Functions*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1 (1967), 183–214, Univ. California Press, Berkeley, CA.

[DueLeu] L. Dümbgen and C. Leuenberger, *Explicit Bounds for the Approximation Error in Benford's Law*, Electronic Communications in Probability **13** (2008), 99–112.

[DurHP] C. Durtschi, W. Hillison and C. Pacini, *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*, Journal of Forensic Accounting, 2004.

[EdH..] X. Edelsbrunner, K. Huan, B. Mackall, S. J. Miller, J. Powell, C. Turnage-Butterbaugh and M. Weinstein, *Benford's Law, the Cauchy Distribution and Financial Data*, preprint 2014.

[Ele1] Elections Canada, *38th General Election – Poll-by-Poll Results – Raw Data*, 2006.
`http://www.elections.ca/scripts/resval/ovr_2004.asp?prov=&lang=e`.

[Ele2] Elections Canada, *39th General Election – Poll-by-Poll Results – Raw Data*, 2006.
`http://www.elections.ca/scripts/resval/ovr_39ge.asp?prov=&lang=e`.

[Ele3] Elections Canada, *40th General Election – Poll-by-Poll Results – Raw Data*, 2010.
`http://www.elections.ca/scripts/resval/ovr_40ge.asp?prov=&lang=e`.

[Ele4] Elections Canada, *41st General Election – Poll-by-Poll Results – Raw Data*, 2012.
`http://www.elections.ca/scripts/resval/ovr_41ge.asp?prov=&lang=e`.

[EngLeu] H. A. Engel and C. Leuenberger, *Benford's Law for Exponential Random Variables*, Statistics & Probability Letters **63** (2003), 361–365.

[EUC1] European Commission, *Eurostat Newsrelease Euroindicators*, Brussels, February 6, 2012.

[EUC2] European Commission, *Communication from the Commission to the European Parliament and the Council, Towards Robust Quality Management for European Statistics*, Brussels, April 15, 2011.

[EUC3] European Commission, *Report on Greek Government Deficit and Debt Statistics*, Brussels, 2010.

[Fan] D. Fanelli, How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data, PLoS ONE 4, no. 5 (2009), 1–11.

[Far] F. Farkas and G. Gyürky, *The Significance of Using the Newcomb-Benford Law as a Test of Nuclear Half-Life Calculations*, Acta Physica Polonica B **41** (2010), no.6, 1213–1221.

[Fef] C. Fefferman, *Pointwise Convergence of Fourier Series*, Ann. of Math. Ser. 2 **98** (1973), 551–571.

[Fel] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd edition, Vol. II, John Wiley & Sons, New York, 1971.

[FerBN] M. J. Ferguson, J. A. Bargh and D. A. Nayak, *After-Affects: How Automatic Evaluations Influence the Interpretation of Subsequent, Unrelated Stimuli*, Journal of Experimental Social Psychology **41** (2005), 182–191.

[FiMil] F. W. K. Firk and S. J. Miller, *Nuclei, Primes and the Random Matrix Connection*, Symmetry **1** (2009), 64–105; doi:10.3390/sym1010064.

[FisVa] M. Fisz and V. S. Varadarajan, *A Condition of Absolute Continuity of Infinitely Divisible Distribution Functions*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **1** (1963), 335–339.

[Fol] G. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd edition, Pure and Applied Mathematics, Wiley-Interscience, New York, 1999.

[Fre] B. S. Frey, *Economists in the PITS?* International Review of Economics **56**, no. 4 (2009), 335–346.

[FriGMH] A. Frigyesi, D. Gisselsson, F. Mitelman and M. Höglund, *Power Law Distribution of Chromosome Aberrations in Cancer*, Cancer Research **63** (2003), 7094–7097.

[Fry] S. Fry, *How Political Rhetoric Contributes to the Stability of Coercive Rule: A Weibull Model of Post-Abuse Government Survival*. Paper presented at the annual meeting of the International Studies Association, Le Centre Sheraton Hotel, Montreal, Quebec, Canada, Mar 17, 2004.

[FuSS] D. Fu, Y. Q. Shi and W. Su, *A Generalized Benford's Law for JPEG Coefficients and its Applications in Image Forensics*, Security, Steganography, and Watermarking of Multimedia Contents IX **6505**, 2007.

[GAISE] GAISE College Report, *Guidelines for Assessment and Instruction in Statistics Education*, American Statistical Association, 2005. `http://www.amstat.org/education/gaise/`.

[GaAl] E. García-Berthou and C. Alcaraz, *Incongruence between Test Statistics and P Values in Medical Papers*, BMC Medical Research Methodology **4** (2004), 13.

[GeWi] C. L. Geyer and P. P. Williamson, *Detecting Fraud in Datasets Using Benford's Law*, Computation in Statistics: Simulation and Computation **33** (2004), no. 1, 229–246.

[GiTo] G. Gigerenzer and P. M. Todd, *Simple Heuristics That Make Us Smart*, ABC Research Group (1999), Oxford University Press, London.

[Gil] D. E. Giles, *Benford's Law and Naturally Occurring Prices in Certain eBay Auctions*, Applied Economics Letters **14** (2007), no. 3, 157–161.

[GoJM] A. Golan, G. Judge and D. Miller, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, John Wiley & Sons, 1996.

[GonPa] J. Gonzales-Garcia and G. Pastor, *Benford's Law and Macroeconomic Data Quality*, International Monetary Fund, Working Paper, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1356437`, 2009.

[GoPa] J. Gonzalez-Garcia and P. Gonzalo, *Benford's Law and Macroeconomic Data Quality*, IMF Working Paper WP/09/10, 2009.

[GoAPA] D. González-Jiménez, E. Argones-Rua, F. Pérez-González and J. L. Alba-Castro, *Benford's Law for Natural and Synthetic Images*, IEEE International Conference on Image Processing (2010), 1245–1248.

[GoWo] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison–Wesley, 1992.

[GrLaW] M. Graber, A. Launov, K. Wälde, *Publish or Perish? The Increasing Importance of Publications for Prospective Economics Professors in Austria, Germany and Switzerland*, German Economic Review **9**, no. 4 (2008), 457–472.

[GrRy] I. S. Gradshteyn and I. M. Ryshik, *Tables of Integrals, Series, and Products*, Academic Press, New York, 1980.

[GrKnP] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics. A Foundation for Computer Science*, 2nd edition, Addison–Wesley, Reading MA, 1994.

[GrHaPa] S. Graham, J. Hasseldine and D. Paton, *Statistical Fraud Detection in a Commercial Lobster Fishery*, New Zealand Journal of Marine and Freshwater Research **43** (2009), no. 1, 457–463.

[GrPa] T. Grammatikos and N. I. Papanikolaou, *Using Benford's Law to Detect Fraudulent Practices in Banking*, SSRN Scholarly Paper No. ID 2352775, 2013, Rochester, NY: Social Science Research Network.

[GrMo] S. Grandison and R. Morris, *Biological Pathway Kinetic Rate Constants Are Scale-Invariant*, Bioinformatics **24** (2008), no. 6, 741–743.

[GrPaFl] R. Gregory, M. Partridge and M. A. Flower, *Performance Evaluation of the Philips Gemini PET/CT System*, IEEE Trans. Nucl. Sci., 53(1), Feb. 2006., 93–101,

[GrJuSc] M. Grendar, G. Judge and L. Schechter, *An Empirical Non-parametric Likelihood Family of Data-Based Benford-Like Distributions*, Physica A: Statistical Mechanics and its Applications, 2006.

[Gsch] T. Gschwend, *Ticket-Splitting and Strategic Voting under Mixed Electoral Rules: Evidence from Germany*, European Journal of Political Research **46** (2007), no. 1, 1–23.

[GüTd] S. Günnel and K.-H. Tödter, *Does Benford's Law Hold in Economic Research and Forecasting?*, Empirica **36** (2009), 273–292.

[HaBTV] E. M. Haacke, R. W. Brown, M. R. Thomas and R. Venkateson, *Magnetic Resonance Imaging Physical Principles and Sequence Design*, Wiley-Liss, New York, USA, 1999.

[Hai] J. Haigh, *The Statistics of National Lottery*, Journal of the Royal Statistical Society, Series A **160** (1997), no. 2, 187–206.

[HalDe] A. R. Halpern and S. D. Devereaux, *Lucky Numbers: Choice Strategies in the Pennsylvania Daily Number Game*, Bulletin of the Psychonomic Society **27** (1989), no. 2, 167–170.

[Ham] D. S. Hamermesh, *Viewpoint: Replication in Economics*, Canadian Journal of Economics **40**, no. 3 (2007), 715–733.

[Ha] R. W. Hamming, *On the distribution of numbers*, Bell Syst. Tech. J. **49** (1970), 1609–1625.

[Han] D. Hand, *Deception and Dishonesty with Data: Fraud in Science*, Significance **4** (2007), 22–25.

[HaWr] G. H. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, 5th edition, Oxford Science Publications, Clarendon Press, Oxford, 1995.

[Has] W. K. Hastings. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, Biometrika **57** (1970), no. 1, 97–109.

[Hay1] S. Hayes, *Does Terminal Digit Preference Occur in Pathology?*, Journal of Clinical Pathology **61** (2008), no. 8, 975–976.

[Hay2] S. Hayes, *Terminal Digit Preference Occurs in Pathology Reporting Irrespective of Patient Management Implication*, Journal of Clinical Pathology **61** (2008), no. 9, 1071–1072.

[Hay3] S. Hayes, *Benford's Law in Relation to Terminal Digit Preference*, Journal of Clinical Pathology **62** (2009), no. 6, 574–575.

[Hej] D. Hejhal, *On a Result of Selberg Concerning Zeros of Linear Combinations of L-Functions*, International Math. Res. Notices **11** (2000), 551–557.

[Hen] N. Henze, *A Statistical and Probabilistic Analysis of Popular Lottery Tickets*, Statistica Neerlandica **51** (1997), no. 2, 155–163.

[HePa] M. Herrmann and F. Pappi, *Stategic Voting in German Constituencies*, Electoral Studies **27** (2008), no. 2, 228–244.

[HeSt] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer, Berlin, 1969.

[Hi1] T. P. Hill, *Random-Number Guessing and the First Digit Phenomenon*, Psychological Reports **6** (1988), 967–971.

[Hi2] T. P. Hill, *Base-Invariance Implies Benford's Law*, Proceedings of the American Mathematical Society **123** (1995), 887–895.

[Hi3] T. Hill, *The First-Digit Phenomenon*, American Scientist **86** (1996), 358–363.

[Hi4] T. Hill, *A Statistical Derivation of the Significant-Digit Law*, Statistical Science **10** (1996), 354–363.

[HiSi] M. Hindry and J. Silverman, *Diophantine Geometry: An Introduction*, Graduate Texts in Mathematics **201**, Springer, New York, 2000.

[HjLR] A. Hjorungnes, J. M. Lervik and T. A. Ramstad, *Entropy Coding of Composite Sources Modeled by Infinite Gaussian Mixture Distributions*, IEEE Workshop on Digital Signal Processing (1996), 235–238.

[HofHo] P. Hofmarcher and K. Hornik, *First Significant Digits and the Credit Derivative Market During the Financial Crisis*, Contemporary Economics **7** (2013), no. 2, 21–29.

[HogMC] R. V. Hogg, J. W. McKean and A. T. Craig, *Introduction to Mathematical Statistics*, 6th edition, Prentice–Hall, Englewood Cliffs, NJ, 2005.

[Hol] P. J. Holewijn, *On the Uniform Distribution of Sequences of Random Variables*, Z. Wahrscheinlichkeitstheorie verw. Geb. **14** (1969), 89–92.

[HoKKS] B. Horn, M. Kreuzer, E. Kochs and G. Schneider, *Different States of Anesthesia Can Be Detected by Benford's Law*, Journal of Neurosurgical Anesthesiology **18** (2006), no. 4, 328–329.

[HoRJB] D. Hoyle, M. Rattray, R. Jupp and A. Brass, *Making Sense of Microarray Data Distributions*, Bioinformatics **18** (2002), no. 4, 576–584.

[Hsü] E. H. Hsü, *An Experimental Study on "Mental Numbers" and a New Application*, Journal of General Psychology **38** (1948), 57–67.

[Hu] W. Hurlimann, *Benford's Law from 1881 to 2006*, http://arxiv.org/pdf/math/0607168.

[Ins1] Instituto Federal Electoral, *Estadísticas y Resultados Electorales*, http://www.ife.org.mx/computos2006/distritales/ReporteDiputadoMR%5B*%5D%5B*%5D.html, where $*$ stands for various numerals (accessed July 13, 2006).

[Ins2] Instituto Federal Electoral, *CG391/2011 Resolución del Consejo General del Instituto Federal Electoral sobre la Solicitud de Registro del Convenio de Coalición Total para Postular Candidato a Presidente de los Estados Unidos Mexicanos, así como Candidatos a Senadores y Diputados por el Principio de Mayoría Relativa, Presentado por los Partidos Políticos Nacionales de la Revolución Democrática, del Trabajo y Movimiento Ciudadano, para Contender en el Proceso Electoral Federal 2011-2012*, http://www.ife.org.mx/docs/IFE-v2/DS/DS-CG/DS-SesionesCG/CG-resoluciones/201%1/noviembre/CGex201111-28/CGe281111rp3.pdf.

[Ins3] Instituto Federal Electoral, *CG73/2012 Resolución del Consejo General del Instituto Federal Electoral sobre la Solicitud de Modificación del Convenio de Coalición Parcial Denominada "Compromiso por México" para Postular Candidatos a Presidente de los Estados Unidos Mexicanos, Senadores y Diputados por el Principio de Mayoría Relativa, Presentada por los Partidos Políticos Nacionales Revolucionario Institucional, Verde Ecologista de México y Nueva Alianza para Contender en el Proceso Electoral Federal 2011-2012*, http://www.ife.org.mx/docs/IFE-v2/DS/DS-CG/DS-SesionesCG/CG-resoluciones/201%2/Febrero/CEex201202-08/CGe80212rp6.pdf.

[Ins4] Instituto Federal Electoral, *Clausulado Integral del Convenio de la Coalición Parcial Denominada "Compromiso por México" Celebrado por el Partido Revolucionario Institucional y el Partido Verde Ecologista de México*, http://www.ife.org.mx/docs/IFE-v2/DEPPP/DEPPP-ConveniosCoalicion/alianzas_D%EPPP/alianzas_DEPPP-pdf/2011-2012/clausula-convenio-compromexico.pdf

[Ins5] Instituto Federal Electoral, *Estadísticas y Resultados Electorales. Descarga de la base de datos de los Cómputos Distritales*, URL: http://www.ife.org.mx/portal/site/ifev2/Estadisticas_y_Resultados_Electorales (accessed October 19, 2012).

[In] J. Interlandi, *An Unwelcome Discovery*, New York Times Magazine, October 2006.

[IRS] Internal Revenue Service, *General Description Booklet for the 1988 Individual Public Use Tax File*, U.S. Department of the Treasury, Statistics of Income Division, 1989. Available at `http://www.nber.org/~taxsim/gdb/`.

[IsKaCh] N. Israeloff, M. Kagalenko and K. Chan, *Can Zipf Distinguish Language From Noise in Noncoding DNA?*, Physical Review Letters **76** (1996), no. 11, 1976.

[Iv] A. Ivić, *On Small Values of the Riemann Zeta-Function on the Critical Line and Gaps between Zeros*, Lietuvos Matematikos Rinkinys **42** (2002), 31–45.

[IwKo] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS, Providence, RI, 2004.

[Iw] H. Iwasawa, *Gaussian Integral Puzzle*, Math. Intelligencer **31** (3) (2009), 38–41.

[JaThYe] M. Jameson, J. Thorner and L. Ye, *Benford's Law for Coefficients of Newforms*, preprint 2014. `http://arxiv.org/abs/1407.1577`.

[JaKKKM] D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of Distributions, Hierarchical Bayesian Models and Benford's Law*, Journal of Algebra, Number Theory: Advances and Applications **1**, no. 1 (March 2009), 37–60.

[JanRu] E. Janvresse and T. de la Rue, *From Uniform Distribution to Benford's Law*, Journal of Applied Probability **41** (2004), no. 4, 1203–1210.

[Jay] E. T. Jaynes, *Information Theory and Statistical Mechanics*, Phys. Rev., 1957.

[JesWi] B. Jessen and A. Wintner, *Distribution Functions and the Riemann Zeta Function*, Transactions of the AMS **38** (1935), 48–88.

[Joa] T. Joachims, *Making Large-Scale SVM Learning Practical*, Advances in Kernel Methods - Support Vector Learning, MIT Press, 1999.

[Jol] J. M. Jolion, *Images and Benford's Law*, Journal of Mathematical Imaging and Vision **14** (2001), 73–81.

[JuSc] G. Judge and L. Schechter, *Detecting Problems in Survey Data Using Benford's Law*, Journal of Human Resources **44** (2009), no. 1, 1–24.

[Jud] H. F. Judson, *The Great Betrayal: Fraud in Science*, Harcourt, Orlando, FL, 2004.

[KaMeb] K. Kalinin and W. Mebane, *Understanding Electoral Frauds through Evolution of Russian Federalism: From "Bargaining Loyalty" to "Signaling Loyalty"*. Paper presented at the 2011 Annual Meeting of the Midwest Political Science Association, Chicago, IL, March 31–April 2, 2011.

[Ka] O. Kallenberg, *Foundations of Modern Probability*, 2nd edition, Springer, Berlin, 2002.

[KaSh] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics **113**, Springer, 1988.

[Kar] A. Karlsson, *Applications of Heat Kernels on Abelian Groups: $\zeta(2n)$, Quadratic Reciprocity, Bessel Integral*, preprint.

[KaSa] N. Katz and P. Sarnak, *Zeros of Zeta Functions and Symmetries*, Bull. AMS **36** (1999), 1–26.

[Kat] Y. Katznelson, *An Introduction to Harmonic Analysis*, 3rd edition, Cambridge University Press, Cambridge, 2004.

[KeSn] J. P. Keating and N. C. Snaith, *Random Matrix Theory and $\zeta(1/2+it)$*, Comm. Math. Phys. **214** (2000), no. 1, 57–89.

[Kemp] J. Kemperman, *Bounds on the Discrepancy Modulo 1 of a Real Random Variable*, Institute of Mathematical Statistics Bulletin **4**, no. 1 (1975), 138.

[Kh] A. Khintchin, *Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze*, Mat. Sbornik **44** (1937), 79–119.

[Ki..] G. King, B. Palmquist, G. Adams, M. Altman, K. Benoit, C. Gay, J. Lewis, R. Mayer and E. Reinhardt, *The Record of American Democracy, 1984–1990*, Harvard University, Cambridge, MA, 1997.

[Kles] J. Klesner, *The July 2006 Presidential and Congressional Elections in Mexico*, Electoral Studies **26** (2007), no. 4, 803–808.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison–Wesley, 3rd edition, 1997.

[Kol] A. N. Kolmogorov, *Une série de Fourier-Lebesgue divergent partout*, Comptes Rendus **183** (1926), 1327–1328.

[KonMi] A. Kontorovich and S. J. Miller, *Benford's Law, Values of L-Functions and the $3x+1$ Problem*, Acta Arith. **120** (2005), 269–297.

[KonSi] A. Kontorovich and Ya. G. Sinai, *Structure Theorem for (d, g, h)-Maps*, Bull. Braz. Math. Soc. (N.S.) **33** (2002), no. 2, 213–224.

[Kosh] T. Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley-Interscience, New York, 2001.

[Koss1] A. E. Kossovsky, *Towards a Better Understanding of the Leading Digits Phenomena*, preprint. http://arxiv.org/abs/math/0612627.

[Koss2] A. E. Kossovsky, *Benford's Law*, World Scientific, 2014.

[Kou] S. Kou, *A Jump-Diffusion Model for Option Pricing*, Management Science **48** (2002), 1086–1101.

[KouWa] S. Kou and H. Wang, *Option Pricing under a Double Exponential Jump-Diffusion Model*, Management Science **50** (2004), 1178–1192.

[Kr] V. M. Kruglov, *A Note on Infinitely Divisible Distributions*, Theory Probab. Appl. **15** (1970), 319–324.

[Kry] J. Krygier, *Psychological Relevance of Benford's Law: Sensitivity to Statistical Regularities and Implications for Theories of Number Representation*, Unpublished Honours thesis (2009), University of Sydney, Australia.

[Kub] M. Kubovy, *Response Availability and the Apparent Spontaneity of Numerical Choices*, Journal of Experimental Psychology: Human Performance and Performance **2** (1977), 359–364.

[KuiNi] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley & Sons, 1974.

[Kul] S. Kullback, *Information Theory and Statistics*, Wiley Publication in Mathematical Statistics, 1959.

[Lad] R. Ladouceur, D. Dube, I. Giroux, N. Legendre and C. Gaudet, *Cognitive Biases and Playing Behavior on American Roulette and the 6/49 Lottery* (1996), unpublished manuscript, Université Laval.

[LafSi] J. M. Lafaille and G. Simonis, *Dissected Re-assembled: An Analysis of Gaming*, 2005.

[Lag1] J. Lagarias, *The $3x+1$ Problem and its Generalizations*, Organic Mathematics (Burnaby, BC, 1995), 305–334, CMS Conf. Proc. **20**, Amer. Math. Soc., Providence, RI, 1997.

[Lag2] J. Lagarias, *The $3x + 1$ Problem: An Annotated Bibliography*, http://arxiv.org/abs/math/0309224.

[Lag3] J. Lagarias, *The Ultimate Challenge: The $3x + 1$ Problem*, American Mathematical Society, Providence, RI, 2010.

[LagSo] J. Lagarias and K. Soundararajan, *Benford's Law for the $3x+1$ Function*, J. London Math. Soc. **74** (2006), ser. 2, no. 2, 289–303.

[LanLi] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd edition, Butterworth-Heinemann, Oxford, 1996.

[Lan] D. Lando, *Credit Risk Modeling*, Princeton University Press, Princeton, NJ, 2004.

[Lang1] S. Lang, *Undergraduate Analysis*, 2nd edition, Springer, New York, 1997.

[Lang2] S. Lang, *Complex Analysis*, Graduate Texts in Mathematics, vol. 103, Springer, New York, 1999.

[Lau] A. Laurinčikas, *Limit Theorems for the Riemann Zeta-Function on the Critical Line II*, Lietuvos Mat. Rinkinys **27** (1987), 489–500.

[LedTa] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer, Berlin, 1991.

[LeDuc1] L. LeDuc, *The Federal Election in Canada, June 2004*, Electoral Studies **24** (2005), no. 2, 338–344.

[LeDuc2] L. LeDuc, *The Federal Election in Canada, January 2006*, Electoral Studies **26** (2007), no. 3, 716–720.

[LeDuc3] L. LeDuc, *The Federal Election in Canada, October 2008*, Electoral Studies **28** (2009), no. 2, 326–329.

[LeDuc4] L. LeDuc, *The Federal Election in Canada, May 2011*, Electoral Studies **31** (2012), no. 1, 239–242.

[LeCoJu] J. Lee, W. Cho and G. Judge, *Stigler's Approach to Recovering the Distribution of First Significant Digits in Natural Data Sets*, Statistics and Probability Letters **80** (2010), no. 2, 82–88.

[LeScEv] L. M. Leemis, B. W. Schmeiser and D. L. Evans, *Survival Distributions Satisfying Benford's Law*, American Statistician **54** (2000), no. 4, 236–241.

[Lév1] P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937.

[Lév2] P. Lévy, *L'addition des variables aléatoires définies sur une circonférence*, Bull. de la S.M.F. **67** (1939), 1–41.

[Ley] E. Ley, *On the Peculiar Distribution of the U.S. Stock Indices Digits*, American Statistician **50** (1996), no. 4, 311–313.

[LiBEM] J. A. List, C. D. Bailey, P. J. Euzent and T. L. Martin, *Academic Economists Behaving Badly? A Survey on Three Areas of Unethical Behavior*, Economic Inquiry **39**, no. 1 (2001), 162–170.

[LoeBr] T. Loetscher and P. Brugger, *Exploring Number Space by Random Digit Generation*, Exp. Brain Res. **180** (2007), 655–665.

[Lop] J. A. G. Lopez, *Fraude Electoral?*, Doble Hélice Ediciones, Chihuahua, Mexico, 2009.

[LoRSST] T. Louwers, B. Ramsey, D. Sinason, J. Strawser and J. Thibodeau, *Auditing & Assurance Services*, 4th edition, Burr Ridge, IL, Irwin–McGraw–Hill, 2011.

[Loy] R. M. Loynes, *Some Results in the Probabilistic Theory of Asympototic Uniform Distributions Modulo* 1, Z. Wahrscheinlichkeitstheorie verw. Geb. **26** (1973), 33–41.

[LePeWi] D. A. Levin, Y. Peres and E. L. Wilmer, *Markov Chains and Mixing Times*, AMS, Providence, RI, 2009.

[Luo] W. Luo, *Zeros of Hecke L-Functions Associated with Cusp Forms*, Acta Arith. **71** (1995), no. 2, 139–158.

[MaRa] N. Madras and D. Randall, *Markov Chain Decomposition for Convergence Rate Analysis*, Ann. Appl. Probab. **12** (2002), no. 2, 581–606.

[Mag] B. Magaloni, *Voting for Autocracy: Hegemonic Party Survival and Its Demise in Mexico*, Cambridge University Press, New York, 2006.

[ManHu] B. Mandelbrot and R. L. Hudson, *The (Mis)behavior of Markets. A Fractal View of Risk, Ruin, and Reward*, Basic Books, New York, 2004.

[MaRRSD] P. Manoochehrnia, F. Rachidi, M. Rubinstein, W. Schulz and G. Diendorfer, *Benford's Law and Its Application to Lightning Data*, IEEE Transactions on Electromagnetic Compatibility **52** (2010), no. 4, 956–961.

[MBGHPSS] R. Mantegna, S. Buldyrev, A. Goldberger, S. Havlin, C. Peng, M. Simons and H. Stanley, *Linguistic Features of Noncoding DNA Sequences*, Physical Review Letters **73** (1994), no. 23, 3169–3172.

[MaAdV] B. C. Martinson, M. S. Anderson and R. de Vries, *Scientists Behaving Badly*, Nature **435** (2005), 737–738.

[McCDo] J. McCann and J. Domínguez, *Mexicans React to Electoral Fraud and Political Corruption: An Assessment of Public Opinion and Voting Behavior*, Electoral Studies **17** (1998), no. 4, 483–503.

[McCFKB] K. C. McCulloch, M. J. Ferguson, C. C. K. Kawada, J. A. Bargh, *Taking a Closer Look: On the Operation of Nonconscious Impression Formation*, Journal of Experimental Social Psychology **44** (2008), 614–623.

[McCMH] B. D. McCullough, K. A. McGeary and T. D. Harrison, *Lessons from the JMCB Archive*, Journal of Money, Credit, and Banking **38**, no. 4, (2006), 1093–1107.

[McCVi] B. D. McCullough and H. D. Vinod, *Verifying the Solution from a Nonlinear Solver: A Case Study*, American Economic Review **93** (2003), 873–892.

[McD] C. McDiarmid, *On the Method of Bounded Differences*, in Surveys in Combinatorics (J. Siemons ed.), 148–188, London Math. Soc. Lecture Notes Series **141**, Cambridge University Press, Cambridge, 1989.

[McLPe] G. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, New York, 2000.

[McSABF] B. McShane, M. Adrian, E. T. Bradlow and P. S. Fader, *Count Models Based on Weibull Interarrival Times*, Journal of Business and Economic Statistics **26** (2006), no. 3, 369–378.

[Meb1] W. Mebane, *Election Forensics: Vote Counts and Benford's Law*, Paper prepared for the 2006 Summer Meeting of the Political Methodology Society, UC-Davis, July 20–22, 2006.

[Meb2] W. R. Mebane, Jr., *Note on the Presidential Election in Iran, June 2009*, (2009),
`http://www.umich.edu/∼wmebane/note29jun2009.pdf`.

[Meb3] W. Mebane, *Election Fraud or Strategic Voting?*, Paper prepared for the 2010 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 22–25, 2010.

[Meb4] W. Mebane, *Fraud in the 2009 Presidential Election in Iran?*, Chance **23** (2010), no. 1, 6–15.

[Meb5] W. Mebane, *Comment on 'Benford's Law and the Detection of Election Fraud'*, Political Analysis **19** (2011), no. 3, 269–272.

[Meb6] W. Mebane, *Second-Digit Tests for Voters' Election Strategies and Election Fraud*, Paper prepared for the 2012 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 11–14, 2012.

[Meb7] W. Mebane, *Election Forensics*, Book MS, 2013.

[MebKa1] W. Mebane and K. Kalinin, *Electoral Falsification in Russia: Complex Diagnostics Selections 2003-2004, 2007-2008* (in Russian), Russian Electoral Review **2** (2009), 57–70.

[MebKa2] W. Mebane and K. Kalinin, *Electoral Fraud in Russia: Vote Counts Analysis Using Second-Digit Mean Tests*, Paper prepared for the 2010 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 22–25, 2010.

[MebKe] W. Mebane and T. Kent, *Second Digit Implications of Voters' Strategies and Mobilizations in the United States during the 2000s*, Paper prepared for the 2013 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 11–14, 2013.

[Mer] R. Merton, *Theory of Rational Option Pricing*, Bell J. Economics and Management Science **4** (1973), 141–183.

[Mer2] R. Merton, *Option Pricing When Underlying Stock Returns Are Discontinuous*, J. Financial Economics **3** (1976), 125–144.

[MeRRTT] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines*, Journal of Chemical Physics **21** (1953), no. 6, 1087–1092.

[Metz] M. Metzker, *Sequencing Technologies - The Next Generation*, Nature Review Genetics **11** (2010), 31–46.

[MiSt] T. Michalski and G. Stoltz, *Do Countries Falsify Economic Data Strategically? Some Evidence That They Might*, Review of Economics and Statistics **95** (2013), no. 2, 591–616.

[Miko] P. G. Mikolaj, *Environmental Applications of the Weibull Distribution Function: Oil Pollution*, Science 2 **176** (1972), no. 4038, 1019–1021.

[Mik] T. Mikosch, *Non-Life Insurance Mathematics*, Springer, 2004.

[Mi1] S. J. Miller, *When the Cramér-Rao Inequality Provides No Information*, Communications in Information and Systems **7** (2007), no. 3, 265–272.

[Mi2] S. J. Miller, *A Derivation of the Pythagorean Won-Loss Formula in Baseball*, Chance Magazine **20** (2007), no. 1, 40–48.

[MiNi1] S. J. Miller and M. Nigrini, *The Modulo* 1 *Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.

[MiNi2] S. J. Miller and M. Nigrini, *Order Statistics and Benford's Law*, International Journal of Mathematics and Mathematical Sciences **2008** (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948.

[MiT-B] S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.

[Mir] T. A. Mir, *The Leading Digit Distribution of the Worldwide Illicit Financial Flows*, preprint 2012, `http://arxiv.org/abs/1201.3432`.

[MiBC] N. Mira, J. Becker and I. Sá-Correia, *Genomic Expression Program Involving the Haa1p-Regulon in Saccharomyces cerevisiae Response to Acetic Acid*, Omics **14** (2010), 587–601.

[MitJuMi] R. C. Mittelhammer, G. Judge and D. J. Miller, *Econometric Foundations*, Cambridge University Press, 2000.

[Monod] J. Monod, *The Growth of Bacterial Cultures*, Annual Review of Microbiology **3** (1949), 371–394.

[Mon] H. Montgomery, *The Pair Correlation of Zeros of the Zeta Function*, Analytic Number Theory, Proc. Sympos. Pure Math. **24**, Amer. Math. Soc., Providence, 1973, 181–193.

[MorSPZ] M. Moret, V. de Senna, M. Pereira and G. Zebende, *Newcomb-Benford Law in Astrophysical Sources*, International Journal of Modern Physics C **17** (2006), no. 11, 1597–1604.

[MorSSZ] M. Moret, V. de Senna, M. Santana and G. Zebende, *Geometric Structural Aspects of Proteins and Newcomb-Benford Law*, International Journal of Modern Physics C **20** (2009), no. 12, 1981–1988.

[Morr] K. E. Morrison,*The Multiplication Game*, Mathematics Magazine **83** (2010), 100–110.

[MosDDK] J. Mosimann, J, Dahlberg, N. Davidian and J. Krueger, *Terminal Digits and the Examination of Questioned Data*, Accountability in Research **9** (2002), no. 2, 75–92.

[MosWE] J. Mosimann, C. Wiseman and R. Edelman, *Data Fabrication: Can People Generate Random Digits?*, Accountability in Research **4** (1995), no. 4, 31–55.

[MyOrSh] M. Myagkov, P. Ordeshook and D. Shaikin, *The Forensics of Election Fraud: With Applications to Russia and Ukraine*, Cambridge University Press, New York, 2009.

[NaWWSRGS] U. Nagalakshmi, Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein and M. Snyder, *The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing*, Science **320** (2008), 1344–1349.

[NaSh] K. Nagasaka and J. S. Shiue, *Benford's Law for Linear Recurrence Sequences*, Tsukuba J. Math. **11** (1987), 341–351.

[New] S. Newcomb, *Note on the Frequency of Use of the Different Digits in Natural Numbers*, Amer. J. Math. **4** (1881), 39–40.

[NiRen] D. Ni and Z. Ren, *Benford's Law and Half-Lives of Unstable Nuclei*, European Physical Journal A **38** (2008), 251–255.

[NiWeRe] D. Ni, L. Wei and Z. Ren, *Benford's Law and β-Decay Half-Lives*, Communications in Theoretical Physics **51** (2009), 713–716.

[Nig1] M. Nigrini, *The Detection of Income Tax Evasion through an Analysis of Digital Frequencies*, PhD thesis, University of Cincinnati, OH, USA, 1992.

[Nig2] M. Nigrini, *Using Digital Frequencies to Detect Fraud*, The White Paper **8** (1994), no. 2, 3–6.

[Nig3] M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*, in Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems (M. Ettredge ed.), University of Kansas, Lawrence, KS, 1996, 69–81.

[Nig4] M. Nigrini, *A Taxpayer Compliance Application of Benford's Law*, Journal of the American Taxation Association, 1996.

[Nig5] M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory **16** (1997), no. 2, 52–67.

[Nig6] M. Nigrini, *Adding Value with Digital Analysis*, Internal Auditor, February, (1999), 21–23.

[Nig7] M. Nigrini, *I've Got Your Number: How a Mathematical Phenomenon Can Help CPAs Uncover Fraud and Other Irregularities*, Journal of Accountancy (May 1999), 79–83.

[Nig8] M. Nigrini, *An Assessment of the Change in the Incidence of Earnings Management Around the Enron-Andersen Episode*, Review of Accounting and Finance **4** (2005), no. 1, 92–110.

[Nig9] M. Nigrini, *Data-Driven Forensic Investigation: Using Microsoft Access and Excel to Detect Fraud and Data Irregularities*, John Wiley & Sons, Hoboken, NJ, 2011.

[NiMi1] M. Nigrini and S. J. Miller, *Benford's Law Applied to Hydrology Data – Results and Relevance to Other Geophysical Data*, Mathematical Geology **39** (2007), no. 5, 469–490.

[NiMi2] M. Nigrini and S. J. Miller, *Data Diagnostics Using Second Order Tests of Benford's Law*, Auditing: A Journal of Practice and Theory **28** (2009), no. 2, 305–324. doi:10.2308/aud.2009.28.2.305.

[NiMit] M. Nigrini and L. Mittermaier, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory **16** (1997), no. 2, 52–67.

[NyM] J. Nye and C. Moul, *The Political Economy of Numbers: On the Application of Benford's Law to International Macroeconomic Statistics*, B.E. Journal of Macroeconomics, 2007.

[OrMNO] M. Orita, A. Moritomo, T. Niimi and K. Ohno, *Use of Benford's Law in Drug Discovery Data*, Drug Discovery Today **15** (2010), 328–331.

[Ost] S. Osterloh, *Accuracy and Properties of Germany Business Cycle Forecasts*, Applied Economics Quarterly **54**, no. 1 (2008), 27–57.

[Ow] A. B. Owen, *Empirical Likelihood*, Chapman and Hall, 2001.

[ÖzBa] G. Özer and B. Babacan, *Benford's Law and Digital Analysis: Application on Turkish Banking Sector*, Business and Economics Research Journal **4** (2013), no. 1, 29–41.

[Pad] T. Padmanabhan, *Physical Significance of Planck Length*, Annals of Physics **165** (1985), no. 1, 38–58.

[Pai] C. Pain, *Benford's Law and Complex Atomic Spectra*, Physical Review E **77** (2008), 012102, doi:10.1103/PhysRevE.77.012102.

[PapPU] A. Papoulis, S.U. Pillai and S. Unnikrishna, *Probability, Random Variables, and Stochastic Processes*, McGraw–Hill, New York, 2002.

[Pat] S. J. Patterson, *An Introduction to the Theory of the Riemann Zeta-Function*, Cambridge Studies in Advanced Mathematics 14, Cambridge University Press, 1995.

[PMDPMLS] S. Paul, D. Mytelka, C. Dunwiddie, C. Persinger, B. Munos, S. Lindborg and A. Schact, *How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge*, Nature Reviews Drug Discovery **9** (2010), 203–214.

[PéAH1] F. Pérez-González, C. T. Abdallah and G. L. Heileman, *A Generalization of Benford's Law and Its Application to Images*, European Control Conference, (2007).

[PéAH2] F. Pérez-González, C. T. Abdallah and G. L. Heileman, *Benford's Law in Image Processing*, IEEE International Conference on Image Processing (2007), 405–408.

[PerTo1] L. Pericchi and D. Torres, *La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela*, Reporte Técnico no-definitivo 2a. versión: Octubre 01, 2004.

[PerTo2] L. Pericchi and D. Torres, *Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela*, Statistical Science **26** (2011), no. 4, 502–516.

[PiTTV] L. Pietronero, E. Tosatti, V. Tosatti and A. Vespignani, *Explaining the Uneven Distribution of Numbers in Nature: The Laws of Benford and Zipf*, Physica A: Statistical Mechanics and its Applications, 2001.

[Pin] R. Pinkham, *On the Distribution of First Significant Digits*, Annals of Mathematical Statistics **32**, no. 4 (1961), 1223–1230.

[PoTh] C. Porter and R. Thomas, *Fluctuations of Nuclear Reaction Widths*, Physical Review **104** (1956), 483–491.

[Q–] M. Quarantelli, K. Berkouk, A. Prinster, B. Landeau, C. Svarer, L. Balkay, B. Alfano, A. Brunetti, J.-C. Baron and M. Salvatore, *Integrated Software for the Analysis of Brain PET/SPECT Studies with Partial-Volume Effect Correction*, J. Nucl. Med. **45**, (2004), 192–201.

[R] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011. `http://www.R-project.org`

[Rai] R. A. Raimi, *The First Digit Problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.

[Ran] D. Randall, *Rapidly Mixing Markov Chains with Applications in Computer Science and Physics*, Computing in Science and Engineering **8** (March 2006), no. 2, 30–41.

[RaEtAl] J. Ranstam, M. Buyse, S. George, S. Evans, N. Geller, B. Scherrer, E. Lesaffre, G. Murray, L. Edler, J. Hutton, T. Colton and P. Lachenbruch, *Fraud in Medical Research: An International Survey of Biostatisticians*, Controlled Clinical Trials **21** (2000), 415–427.

[RapBu] A. Rapoport and D. V. Budescu, *Generation of Random Series in Two-Person Strictly Competitive Games*, Journal of Experimental Psychology: General **121** (1992), 352–363.

[RauGBE] B. Rauch, M. Göttsche, G. Brähler and S. Engel, *Fact and Fiction in EU-Governmental Economic Data*, German Economic Review **12** (2011), no. 2, 243–255.

[RauGBK] B. Rauch, M. Göttsche, G. Brähler and T. Kronfeld, *Deficit versus Social Statistics: Empirical Evidence for the Effectiveness of Benford's Law*, Applied Economics Letters **21** (2014), no. 3, 147–151.

[RauGM] B. Rauch, M. Göttsche and F. El Mouaaouy, *LIBOR Manipulation-Empirical Analysis of Financial Market Benchmarks Using Benford's Law*, 2013. Available at SSRN 2363895.

[ReaCr] T. R. C. Read and N. A. C. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data*, Springer, 1988.

[Rei] R. D. Reiss, *Approximate Distributions of Order Statistics*, Springer, New York, 1989.

[RevYo] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd edition, Springer, Berlin, 1999.

[Rob] H. Robbins, *On the Equidistribution of Sums of Independent Random Variables*, Proc. Amer. Math. Soc. **4** (1953), 786–799.

[RobCa] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2004.

[RobSt] C. J. Roberts and T. D. Stanley, *Meta-Regression Analysis: Issues of Publication Bias in Economics*, Blackwell Publishing, Oxford, UK, 2005.

[Rod] R. Rodriguez, *First Significant Digit Patterns from Mixtures of Uniform Digits*, American Statistician, 2004.

[RorBr] C. Rorden and M. Brett, *Stereotaxic Display of Brain Lesions*, Behav. Neurol. **12** (2000), 191–200.

[Rot] K. Roth, *Rational Approximations to Algebraic Numbers*, Mathematika **2** (1955), 1–20.

[Rou] B. F. Roukema, *A First-Digit Anomaly in the 2009 Iranian Presidential Election*, Journal of Applied Statistics **41** (2014), 164–199. DOI:10.1080/02664763.2013.838664. `http://arXiv.org/abs/0906.2789`.

[RouMA] O. G. Rousset, Y. Ma and A. C. Evans, *Correction for Partial Volume Effects in PET: Principle and Validation*, Journal of Nuclear Medicine **39(5)**, May 1998, 904–911.

[Roy] H. L. Royden, *Real Analysis* 3rd edition, Macmillan, New York, 1988.

[Rud] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, International Series in Pure and Applied Mathematics, McGraw–Hill, New York, 1976.

[Sak] H. Sakamoto, *On the Distributions of the Product and the Quotient of the Independent and Uniformly Distributed Random Variables*, Tôhoku Math. J. **49** (1943), 243–260.

[Sala1] Sala Superior, *Cómputo Final de la Elección Presidencial, Declaración de Validez del Proceso Electoral, y de Presidente Electo de los Estados Unidos Mexicanos*, Tribunal Electoral del Poder Judicial de la Federación, Comisión Calificadora: Magistrados Constancio Carrasco Daza, Flavio Galván Rivera y Salvador Olimpo Nava Gomar, August 31, 2012.

http://www.ife.org.mx/documentos/proceso_2011-2012/documentos/computo_final_calificacion_jurisdiccional.pdf.

[Sala2] Sala Superior, *Juicio de Inconformidad*, Tribunal Electoral del Poder Judicial de la Federación, Expediente: SUP-JIN-359/2012, August 30, 2012, http://www.ife.org.mx/documentos/proceso_2011-2012/documentos/SUP-JIN-359-2012.pdf.

[S-i-M] X. X. Sala-i-Martin, *I Just Ran Two Million Regressions*, American Economic Review **87**, no.2 (1997), 178–183.

[SamTa] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman & Hall, London, 1994.

[Sande] R. Sandels, *Massive Fraud in Mexico's Presidential Elections*, Centre for Research on Globalization, October 15, 2012. http://www.globalresearch.ca/massive-fraud-in-mexicos-presidential-elections/5308401.

[Sando] F. Sandon, *Do Populations Conform to the Law of Anomalous Numbers?*, Population **57** (2002), 753–761.

[Sat1] K.-I. Sato, *Absolute Continuity of Multivariate Distributions of Class* L, J. Multivariate Analysis **12** (1982), 89–94.

[Sat2] K.-I. Sato, *Lévy Processes and Infinite Divisible Distributions*, Cambridge University Press, Cambridge, 2005.

[ScSMG] Ch. Schaefer, J. P. Schrapler, J. P. Muller and G. G. Wagner, *Automatic Identification of Faked and Fraudulent Interviews in the German SOEP*, Journal of Applied Social Science Studies **125** (2005), 183–193.

[SchSti] G. Schaefer and M. Stich, *UCID: An Uncompressed Colour Image Database*, Storage and Retrieval Methods and Applications for Multimedia **5307** (2004), 472–480.

[Sc1] P. Schatte, *On Sums Modulo* 2π *of Independent Random Variables*, Math. Nachr. **110** (1983), 243–261.

[Sc2] P. Schatte, *On the Asymptotic Uniform Distribution of Sums Reduced mod 1*, Math. Nachr. **115** (1984), 275–281.

[Sc3] P. Schatte, *On the Asymptotic Logarithmic Distribution of the Floating-Point Mantissas of Sums*, Math. Nachr. **127** (1986), 7–20.

[SchWil] H. T. Schreuder and M. S. Williams, *Reliability of confidence intervals calculated by bootstrap and classical methods using the IA 1-ha plot design*, in Gen. Tech Rep. RMRS-GTR-57 (2000), Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

[Schum] P. Schumer, *Mathematical Journeys*, Wiley-Interscience, John Wiley & Sons, New York, 2004.

[Schür1] K. Schürger, *Wahrscheinlichkeitstheorie*, Oldenbourg, München, 1998.

[Schür2] K. Schürger, *Extensions of Black-Scholes Processes and Benford's Law*, Stochastic Processes and Their Applications **118** (2008), 1219–1243.

[ScoBM] S. K. Scott, P. J. Barnard and J. May, *Specifying Executive Representations and Processes in Number Generation Tasks*, Quarterly Journal of Experimental Psychology: Section A **54** (2001), 641–664.

[Sel1] A. Selberg, *Contributions to the Theory of the Riemann Zeta-Function*, Arch. Math. Naturvid. **48** (1946), no. 5, 89–155.

[SelBaBe] T. Sellke, M. Bayarri and J. Berger, *Calibration of p Values for Testing Precise Null Hypotheses*, American Statistician **55** (2001), no. 1, 62–71.

[Ser] J. P. Serre, *A Course in Arithmetic*, Springer, 1996.

[Shan] C. E. Shannon, *A Mathematical Theory of Communication*, Bell Technical Journal, 1948.

[ShaMa1] L. Shao and B. Ma, *First Digit Distribution of Hadron Full Width*, Modern Physics Letters A **24** (2009), no. 40, 3275–3282.

[ShaMa2] L. Shao and B. Ma, *Empirical Mantissa Distributions of Pulsars*, Astroparticle Physics **33** (2010), 255–262.

[ShaMa3] L. Shao and B. Ma, *The Significant Digit Law in Statistical Physics*, Physica A **389** (2010), 3109–3116.

[ShaMa4] L. Shao and B. Ma, *First Digit Law in Nonextensive Statistics*, Physical Review E **82** (2010), 041110, doi:10.1103/PhysRevE.82.041110

[ShiHeTh] S. Shikano, M. Herrmann and P. Thurner, *Strategic Voting under Proportional Representation: Threshold Insurance in German Elections*, West European Politics **32** (2009), no. 3, 634–656.

[ShiMa] S. Shikano and V. Mack, *When Does the Second-Digit Benford's Law-Test Signal an Election Fraud? Facts or Misleading Test Results*, Jahrbücher fur Nationalökonomie und Statistik **231** (2009), no. 5–6, 719–732.

[Shi] A. N. Shiryayev, *Probability*, Springer, Berlin, 1984.

[Sim] J. Simon, *An Analysis of the Distribution of Combinations Chosen by UK National Lottery Players*, Journal of Risk and Uncertainty **17** (1999), no. 3, 243–276.

[Sin] Ya. G. Sinai, *Statistical $(3x + 1)$ Problem*, Comm. Pure Appl. Math. **56** (2003), no. 7, 1016–1028.

[Sinc1] A. J. Sinclair, *Algorithms for Random Generation and Counting. A Markov Chain Approach*, in Progress in Theoretical Computer Science, Birkhäuser, Boston, MA, 1993.

[Sinc2] A. J. Sinclair, *Convergence Rates for Monte Carlo Experiments*, in Numerical Methods for Polymeric Systems (Minneapolis, MN, 1996), 1–17 (S. G. Whittington ed.), IMA Vol. Math. Appl. **102**, Springer, New York, 1998.

[Sm] S. W. Smith, *Explaining Benford's Law*, Chapter 34 in The Scientist and Engineer's Guide to Digital Signal Processing, 1997. Republished in softcover by Newnes, 2002.

[Sor] D. Sornette, *Critical Phenomena in Natural Sciences*, Springer, 2000.

[Sta] W. H. Starbuck, *How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication*, Organization Science **16**, no. 2 (2005), 180–200.

[StSh1] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, Princeton, NJ, 2003.

[StSh2] E. Stein and R. Shakarchi, *Complex Analysis*, Princeton University Press, Princeton, NJ, 2003.

[SS3] E. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, Princeton, NJ, 2005.

[SpTh] M. D. Springer and W. E. Thompson, *The Distribution of Products of Independent Random Variables*, SIAM J. Appl. Math. **14** (1966), 511–526.

[Sti] G. Stigler, *The Distribution of Leading Digits in Statistical Tables*. Note, written 1945–1946, Stigler's 1975 address was Haskell Hall, University of Chicago, Chicago, IL 60637.

[Str] K. Stromberg, *Probabilities on a Compact Group*, Trans. Amer. Math. Soc. **94** (1960), 295–309.

[Sug] G. Sugihara, *Minimal Community Structure: An Explanation of Species Abundance Patterns*, American Naturalist **116** (1980), no. 6, 770–787.

[TeKaDu] Y. Terawaki, T. Katsumi and V. Ducrocq, *Development of a Survival Model with Piecewise Weibull Baselines for the Analysis of Length of Productive Life of Holstein Cows in Japan*, Journal of Dairy Science **89** (2006), no. 10, 4058–4065.

[Ti] H. Tijms, *Understanding Probability: Chance Rules in Everyday Lives*, Cambridge University Press, 2007.

[Töd] K.-H. Tödter, *Benford's Law as an Indicator of Fraud in Economics*, German Economic Review **10**, no. 3 (2009), 339–351.

[TorFGS] J. Torres, S. Fernández, A. Gamero and A. Sola, *How Do Numbers Begin? (The First Digit Law)*, European Journal of Physics **28** (2007), L17–L25.

[Tuc] H. G. Tucker, *Absolute Continuity of Infinitely Divisible Distributions*, Pac. J. Math **12** (1962), 1125–1129.

[Tur] P. R. Turner, *The Distribution of Leading Significant Digits*, IMA J. Numer. Anal. **2** (1982), no. 4, 407–412.

[Va] H. Varian, *Benford's Law*, American Statistician **23** (1972), 65–66.

[Vos] R. Voss, *Comment on "Linguistic Features of Noncoding DNA Sequences"*, Physical Review Letters **76** (1996), no. 11, 1978.

[VoxGo] W. Voxman and R. Goetschel, Jr., *Advanced Calculus*, Mercer Dekker, New York, 1981.

[Wag] S. Wagon, *Benford's Law and Data Spread*, `http://demonstrations.wolfram.com/BenfordsLawAndDataSpread/`, Wolfram Demonstrations Project, Published: March 10, 2009.

[Wall] G. K. Wallace, *The JPEG Still Picture Compression Standard*, Communications of the ACM **34** (1991).

[Walt] W. Walter, *Ordinary Differential Equations*, Springer, 1998.

[Wash] L. Washington, *Benford's Law for Fibonacci and Lucas Numbers*, Fibonacci Quart. **19** (1981), no. 2, 175–177.

[We] W. Weibull, *A Statistical Distribution Function of Wide Applicability*, J. Appl. Mech. **18** (1951), 293–297.

[WCPBKO] K. Wells, J. Chiverton, M. Partridge, M. Barry, H. Kadhem and B. Ott, *Quantifying the Partial Volume Effect in PET using Benford's Law*, IEEE Trans. Nucl. Sci. **54(5)**, Oct. 2007, 1616–1625.

[WerAa] M. Wernick and J. Aarsvold, *Emission Tomography, The Fundamentals of PET and SPECT*, Elsevier Academic Press, 2004.

[Whi] D. Whitford, *Proteins: Structure and Function*, John Wiley & Sons, Chichester, 2005.

[Wig1] E. Wigner, *Results and Theory of Resonance Absorption*. Gatlinburg Conference on Neutron Physics by Time-of-Flight, Oak Ridge, National Lab. Report No. ORNL-2309, 1957; Elsevier Science B.V, 2001.

[Wig2] E. Wigner, *On the Statistical Distribution of the Widths and Spacings of Nuclear Resonance Levels*, Mathematical Proceedings of the Cambridge Philosophical Society **47** (1951), 790–798.

[Wis] J. Wishart, *The Generalized Product Moment Distribution in Samples from a Normal Multivariate Population*, Biometrika **20A** (1928), 32–52.

[Wój] M. R. Wójcik, *A Characterization of Benford's Law through Generalized Scale-Invariance*, Mathematical Social Sciences **71** (2014), 1–5.

[WolAu] H. Wolff and M. Auffhammer, *Endogenous Choice of Development Indicators: Are Development Countries Misclassified? Evidence from the HDI*, Agricultural and Resource Economics, UC Berkeley (2006).

[ZiBGS] W. T. Ziemba, S. L. Brumelle, A. Gautier and S. L. Schwartz, *Dr. Z's 6/49 Lotto Guidebook*, Vancouver and Los Angeles: Dr. Z. Investments., 1986.

[Zip] G. K. Zipf, *Human Behavior and the Principle of Least Effort*, Addison–Wesley Press, 1949.

[ZwJRR] M. Zwietering, I. Jongenburger, F. Rombouts and K. van't Riet, *Modeling of the Bacterial Growth Curve*, Applied and Environmental Microbiology **56** (1990), 1875–1881.

[Zy] A. Zygmund, *Trigonometrical Series*, vols. I and II, Cambridge University Press, Cambridge, 1968.

Index