Graphing conformity of distributions to Benford's Law for various bases

Master's thesis: Jesse Dorrestijn,
jessedorrestijn@gmail.com

Mathematical Sciences, Universiteit Utrecht,

Advisor: Dr. Karma Dajani (Universiteit Utrecht)

October 1, 2008

Contents

1 NOTATION **3**

2 INTRODUCTION **4**

3 BENFORD'S LAW **5**

4 PROBABILITY DISTRIBUTIONS **6**

5 THE MANTISSA DISTRIBUTION **7**

6 TAKE LOGARITHMS AND CALCULATE MODULO ONE **9**
6.1 THE STANDARD NORMAL DISTRIBUTION 15
6.2 THE BETA DISTRIBUTION 17

7 SCALAR MULTIPLICATION **18**

8 A BENFORD-DISTANCE ON DISTRIBUTIONS **22**
8.1 THE NORMAL DISTRIBUTION 27
8.2 THE GAMMA DISTRIBUTION 28
8.3 COMBINATIONS OF DISTRIBUTIONS 29

9 MANTISSA DISTRIBUTION IN SEVERAL BASES **31**
9.1 SAMPLING FUNCTIONS 34
9.2 CONTINUITY OF THE DISTRIBUTION OF THE LOGARITHMS OF THE MANTISSAE 37
9.3 BENFORD TOWARDS BASE 1 38

10 A FUNCTION ON DISTRIBUTIONS **43**

11 ANOTHER FUNCTION ON DISTRIBUTIONS **48**

12 CONCLUSION **51**

13 DANKWOORD (WORDS OF THANKS IN DUTCH) **53**

1 Notation

a.e.	almost everywhere with respect to Lebesgue measure
α_b	a function from $\mathcal{P}^+_{a.c.}$ to $\mathbb{R}^+$, see Definition 11.2
b	a real number bigger than one, called "the base"
β_b	a function from $\mathcal{P}^+_{a.c.}$ to $\mathbb{R}^+$, see Definition 10.1
$\mathcal{B}(\mathbb{R})$	the set of Borel sets on $\mathbb{R}$
f_{BL_b}	the density function corresponding to Benford's Law relative to base b.
F_{BL_b}	the distribution function corresponding to Benford's Law relative to base b
γ_b	a Benford-distance from $\mathcal{P}^+_{a.c.}$ to $[0,2)$, see Definition 8.1
λ	the Lebesgue measure on the measurable space $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R}\Big)$
$m_b(x)$	the mantissa of x relative to base b
$\mathcal{P}$	the class of probability measures on the measurable space $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R})\Big)$
$\mathcal{P}^+$	the subclass of $\mathcal{P}$ consisting of probability measures that give measure zero to the set $(-\infty, 0]$
$\mathcal{P}_{a.c.}$	the subclass of $\mathcal{P}$, such that the probability measures are absolutely continuous with respect to Lebesgue measure
$\mathcal{P}^+_{a.c.}$	the subclass of $\mathcal{P}^+$, such that the probability measures are absolutely continuous with respect to Lebesgue measure
P_X	an element of $\mathcal{P}$ with corresponding random variable X
P_{BL_b}	the probability measure called Benford's Law relative to base b
$\mathbb{P}$	a probability measure on the measurable space $\Big(\Omega, \mathcal{F}\Big)$
$\mathbb{R}^+$	all real numbers equal to or bigger than zero

2 Introduction

At first sight it does not seem a task for mathematicians to investigate first digits of numbers, it has a flavour of amateurism, because the first digit of a number depends on the numeral system being used to represent the number. If one nevertheless investigates numbers obtained from the most various sources, one sees to his surprise that in general the first digits do not follow the expected uniform distribution. The first one to notice this was the mathematician and astronomer Simon Newcomb [NEW] who wrote the following in 1881:

> 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.

Consulting logarithmic tables, which one could use to multiply numbers, Newcomb noticed the strange tendency of numbers to start with a 1. After that he determined that the digits were distributed logarithmically, which enabled him to calculate the frequencies.
More than a half century later, in 1937, Frank Benford [BEN] rediscovered the same law after investigating more than 20000 numbers from various sources. The first digits of surfaces of rivers, street addresses and populations turn out to be distributed approximately logarithmically. That means that more than thirty percent of these numbers starts with a 1. For numbers gotten from front pages of newspapers, numbers which find their origin in a lot of different sources, this approximation is even better. On the contrary, for other numbers like IQ-test results this approximation is very bad.
Behind this strange phenomenon hides a law that appears the moment one writes down numbers in a number system. To write down numbers we use a decimal number system. This predilection for ten is not a coincidence, man has ten fingers. Of course it is possible to denote numbers using another base, for example the Maya Indians were used to denote numbers in base 20 and computers calculate in a binary system.
Benford already noticed on his final page that the phenomenon turns up not only using base 10, but also when using other bases. He wrote:

> ...the logarithmic relationship is not a result of the particular numerical system, with its base 10, that we have elected to use. Any other base, such as 8, or 12, or 20, to select some of the numbers that have been suggested at various times, would lead to similar relationships...

Inspired by Benford and also by present-day mathematicians as Ted Hill and Peter Schatte, in this masters thesis Benford's Law will be examined without clinging to base 10. Usually tests on the degree to which numbers satisfy Benford's Law are only done with respect to base 10. If the frequencies of numbers are equal to the by Benford predicted values in this base, then the numbers are said to satisfy Benford's Law. We will see that statements about the degree to which numbers satisfy Benford's Law should be made with care. In some cases this degree depends on the base being used. It is possible that numbers of which the first-digits are distributed logarithmically with respect to base 10, are not distributed logarithmically with respect to base 20. In this case without mentioning the base a Dutchman, who is counting in base 10, will claim that these numbers follow Benford's Law while a Maya, who is counting in base 20, would never do this. The Dutchman should formulate his claim as: these numbers satisfy Benford's Law with respect to base 10. The same problem arises when only a finite number of significant digits are regarded, in that case the distribution of the digits can depend on the measuring units being used. If for example only the first significant digit is regarded, it is possible that a Dutchman and an American do not agree whether or not their frequencies are equal to the by Benford predicted values.

3 Benford's Law

Originally, investigating numbers concerning Benford's Law is investigating frequencies of first significant digits of numbers. In this thesis Benford's Law is studied by studying mantissae of numbers. Let us give a definition of a mantissa of a real number:

Definition 3.1 Let b be a real number bigger than 1, which we will call a (real) base. For $x \in \mathbb{R}\backslash\{0\}$ define the **mantissa** of x relative to base b, denoted by $m_b(x)$, as the unique $m \in [1,b)$ such that $m \cdot b^k = |x|$ with $k \in \mathbb{Z}$ and define $m_b(0) = 0$. The function $m_b : \mathbb{R} \to \mathbb{R}$ that assigns to x the mantissa $m_b(x)$ is called the **mantissa function** relative to base b.

For example $m_{10}(\pi) = \pi = 3.14\ldots$ and $m_{10}(0.00123) = m_{10}(123) = 1.23$ and $m_{\sqrt{2}}(\pi) = \frac{\pi}{\sqrt{2}^3} = 1.11072\ldots$. Defined in this way the mantissa of a number unequal to zero lies in the interval $[1,b)$. This interval is exactly the interval where Benford's Law lives, more formally: only to subsets of this interval can Benford's Law give strictly positive measure. Let us give the definition of Benford's Law:

Definition 3.2 Let $b > 1$ be a real base. **Benford's Law relative to base** b is the probability measure P_{BL_b} on $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R})\Big)$ determined by the following distribution function:

$$F_{\mathrm{BL}_b}(x) := \begin{cases} 0 & \text{for } x < 1, \\ \log_b(x) & \text{for } x \in [1,b), \\ 1 & \text{for } x \geq b. \end{cases}$$

Important to notice is that for different bases the probability measures with the name Benford's Law are different. For every $b > 1$ Benford's Law is a logarithmic distribution which assigns measure 1 to the interval $[1,b)$. If we take an interval $[c,d] \subset [1,b)$ we have

$$P_{\mathrm{BL}_b}([c,d]) = \log_b(d) - \log_b(c).$$

If mantissae of numbers written down in base $b = 10$ follow Benford's distribution, then the probability that the first significant digit is i, where $i \in \{1,2,\ldots,9\}$, is

$$\log_{10}(i+1) - \log_{10}(i) = \log_{10}(1 + \frac{1}{i}).$$

Mantissae of numbers do often follow Benford's Law. How can we investigate this mathematically? Instead of investigating numbers we will assume numbers to be distributed as some known distribution on $\mathbb{R}$ and look how close the corresponding distribution of the mantissae is to Benford's distribution. In an article of Ted Hill entitled *A Statistical Derivation of the Significant-Digit Law*, published in 1995, is written:

> An interesting open problem is to determine which common distributions (or mixtures thereof) satisfy Benford's Law, i.e., are scale or base-invariant or which have mantissas with logarithmic distributions. For example, the standard Cauchy distribution is close to satisfying Benford's Law (cf. Raimi (1976)), and the standard Gaussian is not, but perhaps certain natural mixtures of some common distributions are.

This open problem gives rise to the following questions, to which we would like to give answers at the end of the thesis:

1. *How should we order distributions on the degree to which they satisfy Benford's Law relative to all bases $b > 1$?*
2. *What properties of a distribution do have great influence on this degree?*

3. *Why do mantissae of numbers often follow Benford's Law?*

In Definition 8.1 we will define a Benford-distance on probability distributions that measures how close the corresponding mantissa distribution relative to **one** base is to Benford's Law. This distance will be zero if and only if the distribution satisfies Benford's Law in this base.

In order to define a distance on probability distributions, we must first define the class of probability distributions we will work with.

4 Probability distributions

Definition 4.1 Let $\mathcal{P}$ be the class of probability distributions P which are probability measures on the measurable space $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R})\Big)$. Let $\mathcal{P}_{a.c.}$ be the subclass in which the P are absolutely continuous with respect to λ, the Lebesgue measure on $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R})\Big)$.

Definition 4.2 Let $\mathcal{P}^+$ be the subclass of $\mathcal{P}$ which consists of probability measures P which give measure 0 to the set $(-\infty, 0]$. Let $\mathcal{P}^+_{a.c.}$ be the subclass of $\mathcal{P}^+$ in which the P are absolutely continuous with respect to λ, the Lebesgue measure on $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R})\Big)$.

The classes $\mathcal{P}$ and $\mathcal{P}_{a.c.}$ are defined in such a way that they contain common distributions like the normal distribution, the exponential distribution, the uniform distribution and Benford's distribution. Later in Section 6 we will see why we defined the classes $\mathcal{P}^+$ and $\mathcal{P}^+_{a.c.}$ and in Theorem 6.3 we will see how these distributions can be adapted such that they become elements of $\mathcal{P}^+$ and $\mathcal{P}^+_{a.c.}$. Let us determine more exactly what kind of distributions are in our classes. A probability measure $P \in \mathcal{P}$ can be determined in several ways. We follow the definitions of D.Williams [WIL].

Theorem 4.3

(I) Let $X : \Omega \to \mathbb{R}$ be a **random variable** defined on the probability space $\Big(\Omega, \mathcal{F}, \mathbb{P}\Big)$, i.e. a $\mathcal{F}$-measurable function. Then $P_X : \mathcal{B}(\mathbb{R}) \to [0,1]$, defined by

$$P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) \quad \text{for all } B \in \mathcal{B}(\mathbb{R}),$$

is a probability measure in $\mathcal{P}$.
The **distribution function** $F_X : \mathbb{R} \to [0,1]$ of P_X defined by

$$F_X(x) = P_X((-\infty, x]) \qquad \text{for all } x \in \mathbb{R},$$

is non-decreasing, right-continuous and furthermore $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.

(II) Let $F : \mathbb{R} \to [0,1]$ be a non-decreasing, right-continuous function with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. Then $P : \mathcal{B}(\mathbb{R}) \to [0,1]$ determined by

$$P((-\infty, x]) = F(x) \quad x \in \mathbb{R},$$

is a probability measure in $\mathcal{P}$. Given this function F there exists a probability space $\Big(\Omega', \mathcal{F}', \mathbb{P}'\Big)$ and a random variable $X : \Omega' \to \mathbb{R}$ such that P is the probability measure defined by X as in (I) of this theorem. We can write $P = P_X$. Furthermore F is the distribution function of P_X. We can write $F = F_X$.

(III) Let $f : \mathbb{R} \to \mathbb{R}$ be a non-negative λ-integrable Borel function with $\int_{\mathbb{R}} f d\lambda = 1$. Then $P : \mathcal{B}(\mathbb{R}) \to [0,1]$ defined by

$$P(B) = \int_B f d\lambda, \quad \forall B \in \mathcal{B}(\mathbb{R})$$

is a probability measure in $\mathcal{P}_{a.c.}$. Furthermore, there exists a probability space $\left(\Omega', \mathcal{F}', \mathbb{P}'\right)$ and a random variable $X : \Omega' \to \mathbb{R}$ with corresponding probability measure $P_X = P$. The function $f_X = f$ is called the **density function** of P_X and is unique almost everywhere.

We have seen that random variables, distribution functions and density functions can define probability measures in $\mathcal{P}$. Conversely, given a probability measure $P \in \mathcal{P}$, there exists a random variable X defined on some probability space such that $P_X = P$. Furthermore, if $P_X \in \mathcal{P}_{a.c.}$, then it has a density function f_X.

Proof: See [WIL] page 33 to 35 and page 68.

5 The mantissa distribution

We want to define what we mean by a (probability) distribution satisfying Benford's Law in base b. First we will define the mantissa distribution, the mantissa distribution function and the mantissa density function. Let $P_X \in \mathcal{P}$ be a probability distribution, with a corresponding random variable X. Then the function $m_b(X)$ is also a random variable. Namely, X is a random variable from Ω to $\mathbb{R}$ and it is $\mathcal{F}$-measurable. The mantissa function $m_b(x)$ from $\mathbb{R}$ to $\mathbb{R}$ is a Borel function. Then the composition lemma for measurable functions says that $m_b(X)$ is a $\mathcal{F}$-measurable function from Ω to $\mathbb{R}$. We conclude that it is a random variable. Theorem 4.3 (I) enables us to define the mantissa distribution, $P_{m_b(X)}$, which is an element of $\mathcal{P}$.

Definition 5.1 Given a probability distribution $P_X \in \mathcal{P}$, the **mantissa distribution** (relative to base b) is the probability measure $P_{m_b(X)} \in \mathcal{P}$. According to Theorem 4.3, it is defined as

$$P_{m_b(X)}(B) = \mathbb{P}((m_b(X))^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega : m_b(X(\omega)) \in B\}) \text{ for all } B \in \mathcal{B}(\mathbb{R}).$$

The distribution function $F_{m_b(X)}$ is called the **mantissa distribution function** (relative to base b). Also according to Theorem 4.3, it is defined as the following function from $\mathbb{R}$ to $[0,1]$

$$F_{m_b(X)}(x) = P_{m_b(X)}((-\infty, x]) = \mathbb{P}(m_b(X) \leq x) = \mathbb{P}(\{\omega \in \Omega : m_b(X(\omega)) \leq x\}).$$

If $P_{m_b(X)}$ is absolutely continuous with respect to λ, then its density function $f_{m_b(X)}$ is called the **mantissa density function** (relative to base b). According to Theorem 4.3 this is a function from $\mathbb{R}$ to $\mathbb{R}$ such that

$$P_{m_b(X)}(B) = \int_B f_{m_b(X)} d\lambda, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Remark that from this definition it follows that the mantissa density function is determined almost everywhere.

In Section 3 in the citation of T.Hill, has been written about distributions satisfying Benford's Law. When does a distribution satisfy Benford's Law?

Definition 5.2 A probability distribution or probability measure $P_X \in \mathcal{P}$ **satisfies Benford's Law in base** b if and only if

- the mantissa distribution relative to base b is equal to Benford's Law relative to base b, that is

$$P_{m_b(X)}(B) = P_{\mathrm{BL}_b}(B) \qquad \forall B \in \mathcal{B}(\mathbb{R}),$$

  or

- the mantissa distribution function relative to base b is equal to the distribution function determining Benford's Law relative to base b. We can write this as

$$F_{m_b(X)}(x) = \left\{ \begin{array}{ll} 0 & \text{for } x < 1, \\ \log_b(x) & \text{for } x \in [1, b), \\ 1 & \text{for } x \geq b. \end{array} \right\} = F_{\mathrm{BL}_b}(x).$$

Benford's Law is absolutely continuous with respect to λ, so equivalent is that

- there is a mantissa density function relative to base b such that

$$f_{m_b(X)}(x) = \left\{ \begin{array}{cc} \frac{1}{x \cdot \log_e(b)} & \text{for a.e. } x \in [1,b), \\ 0 & \text{for a.e. other } x. \end{array} \right\} =: f_{\mathrm{BL}_b}(x).$$

Observe that we always mention the base b. One can not talk about a distribution satisfying Benford's Law without mentioning a base. Later in some examples we will see distributions for which the degree to which they satisfy Benford's Law differs a lot by varying the base.

Let $b > 1$ be an integral base and let $P_X \in \mathcal{P}$. From the mantissa distribution $P_{m_b(X)}$ one can easily derive the expected relative frequencies of the first significant digits of numbers that follow P_X. Namely, the expected frequency of the first significant digit i, where $i \in \{1,2,\ldots,b\}$, is given by

$$P_{m_b(X)}([i,i+1)) = \mathbb{P}\Big(\{\omega \in \Omega : m_b(X(\omega)) \in [i,i+1)\}\Big).$$

Usually one looks at the case that b is equal to 10, as we did after definition 3.2. Following Hill [HI2], if a distribution $P_X \in \mathcal{P}$ satisfies Benford's Law in base 10, then the relative frequency of the first significant digit i, where $i \in \{1,\ldots,9\}$, is given by:

$$P_{m_{10}(X)}([i,i+1)) = \int_{[i,i+1)} f_{m_{10}(X)} d\lambda = \int_{[i,i+1)} \frac{1}{x \cdot \log_e(10)} d\lambda(x) =$$

$$\log_{10}(i+1) - \log_{10}(i) = \log_{10}(1+\frac{1}{i}).$$

If one only considers the second significant digit, then the relative frequency of the digit $i \in \{0,1,\ldots,9\}$ is equal to the following sum of integrals, which we also work out:

$$\sum_{k=1}^{k=9} \int_{k+\frac{i}{10}}^{k+\frac{i+1}{10}} \frac{1}{x \cdot \log_e(10)} d\lambda(x) = \sum_{k=1}^{9} \log_{10}(k+\frac{i+1}{10}) - \log_{10}(k+\frac{i}{10}) =$$

$$\sum_{k=1}^{9} \log_{10}(1+\frac{1}{10 \cdot k + i})$$

Finally, the relative frequency of numbers of the form $d_1.d_2\ldots d_n\ldots$ is equal to the following integral, which we also work out:

$$\int_{d_1.d_2\ldots d_n}^{d_1.d_2\ldots d_n+10^{1-n}} \frac{1}{x \cdot \log_e(10)} d\lambda(x) = \log_{10}(d_1.d_2\ldots d_n + 10^{1-n}) - \log_{10}(d_1.d_2\ldots d_n) =$$

$$\log_{10}\Big(1+\Big(\sum_{j=1}^{j=n} d_j \cdot 10^{n-j}\Big)^{-1}\Big).$$

For instance the relative frequency of mantissae starting with 1.2345 is equal to $\log_{10}(1.2346) - \log_{10}(1.2345)$, or $\log_{10}(1+(12345)^{-1})$.

Already in 1880 Simon Newcomb [NEW] was the first to publish the frequencies of the first and second significant digit, see Table 1.

Of course there is a relation between a given probability distribution and its corresponding mantissa distributions. In the next section we will find an explicit formula which represents this relation.

Dig.		First Digit.	Second Digit.
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: Newcomb's table: The relative frequencies of the first and second significant digit of numbers following a distribution that satisfies Benford's Law in base 10.

6 Take logarithms and calculate modulo one

Given a probability distribution P_X we are interested in $P_{m_b(X)}$. To find a formula relating these distributions it is important to know very well what the mantissa function is. It is difficult to find the formula using Definition 3.1. Therefore, we will give another definition of a mantissa which is equivalent to Definition 3.1, at least for strictly positive numbers. Lemma 6.2 below allows us to give the following definition.

Definition 6.1 Define the mantissa of $x \in \mathbb{R}^+ \backslash \{0\}$ relative to base b as $b^{\log_b(x) (\text{mod } 1)}$.

This definition is easier to use when calculating mantissae of numbers with the computer, and it helps us finding our desired formula.

Lemma 6.2 Let $x \in \mathbb{R}^+ \backslash \{0\}$, then $m_b(x) = b^{\log_b(x) (\text{mod } 1)}$

Proof Since $m_b(x) \in [1, b)$ we have that $\log_b(m_b(x)) \in [0, 1)$, so we can write:

$$b^{\log_b(x) (\text{mod } 1)} = b^{\log_b(m_b(x) \cdot b^k) (\text{mod } 1)} = b^{(\log_b(m_b(x)) + \log_b(b^k)) (\text{mod } 1)} =$$

$$b^{(\log_b(m_b(x)) + k) (\text{mod } 1)} = b^{\log_b(m_b(x)) (\text{mod } 1)} =$$

$$b^{\log_b(m_b(x))} = m_b(x).$$

□

This definition of a mantissa is only equivalent to Definition 3.1 for strictly positive real numbers, because the logarithm is not defined for negative numbers nor for zero. We want to use this definition of a mantissa, so we will look at probability distributions of the class $\mathcal{P}^+$. We have to find a way of converting a probability distribution from $P_X \in \mathcal{P}$ to an element of this class.

Let us see how we can convert a probability distribution of $\mathcal{P}$ to a probability distribution of $\mathcal{P}^+$. The easiest method is taking absolute values. Let $P_X \in \mathcal{P}$. If $P_X(\{0\}) = 0$, then the probability distribution $P_{|X|}$ is in $\mathcal{P}^+$. From now on we will only consider distributions for which $P_X(\{0\}) = 0$. Furthermore, without loss of generality, we assume that $X(\omega) \neq 0$ for all $\omega \in \Omega$. This enables us to take logarithms of the absolute values of the random variables.

The goal of this chapter is to derive a formula which expresses the mantissa distribution function $F_{m_b(X)}$ of a given probability distribution, which is in $\mathcal{P}^+$, in terms of the corresponding distribution function F_X. And even more important for this thesis we want to derive a formula which expresses the mantissa density function $f_{m_b(X)}$ of a probability measure in $\mathcal{P}^+_{a.c.}$ in terms of f_X. We will first look what happens to the distribution function and density function of a given probability distribution $P_X \in \mathcal{P}$ if we take absolute values of X.

Theorem 6.3 Let $P_X \in \mathcal{P}$. Then $P_{|X|} \in \mathcal{P}^+$ and the distribution function of $P_{|X|}$ is given by:

$$F_{|X|}(x) = \begin{cases} F_X(x) - F_X^-(-x) & \text{for } x > 0 \text{ and} \\ 0 & \text{for } x \leq 0. \end{cases}$$

Let P_X be in $\mathcal{P}_{a.c.}$, then $P_{|X|} \in \mathcal{P}^+_{a.c.}$ and a density function of $P_{|X|}$ is given by

$$f_{|X|}(x) = \begin{cases} f_X(x) + f_X(-x) & \text{for a.e. } x > 0 \text{ and} \\ 0 & \text{for a.e. } x \leq 0. \end{cases}$$

Further, the mantissa distribution of $P_{|X|}$ is equal to the mantissa distribution of P_X.

Note that $F_X^-(-x)$ is the left limit of F_X in the point $-x$.

Proof Note that $|X|$ is a random variable. Theorem 4.3 implies that $P_{|X|}$ defined by

$$P_{|X|}(B) = \mathbb{P}(|X| \in B) \qquad \text{for all } B \in \mathcal{B}(\mathbb{R}),$$

is a probability measure, that is $P_{|X|} \in \mathcal{P}$. The distribution function is $F_{|X|}(x) = P_{|X|}((-\infty, x])$. If $x > 0$, then

$$\mathbb{P}(|X| \leq x) = \mathbb{P}(-x \leq X \leq x) = F_X(x) - F_X^-(-x)$$

and if $x \leq 0$, then $P_{|X|}((-\infty, x]) = 0$, because $|X|$ is positive. The positivity of $|X|$ also implies that $P_{|X|} \in \mathcal{P}^+$.
Now let P_X be in $\mathcal{P}_{a.c.}$. Assume first that $x > 0$. Then

$$P_{|X|}((-\infty, x]) = \mathbb{P}(|X| \leq x) = \mathbb{P}(-x \leq X \leq x) =$$

$$\int_{[-x,x]} f_X d\lambda = \int_{[0,x]} f_X d\lambda + \int_{[-x,0]} f_X d\lambda = \int_{[0,x]} f_X d\lambda + \int_{[0,x]} f_{-X} d\lambda = \int_{[0,x]} f_X + f_{-X} d\lambda =$$

$$\int_{(-\infty,x]} 1_{[0,\infty)} \cdot (f_X + f_{-X}) d\lambda = \int_{(-\infty,x]} 1_{[0,\infty)} \cdot (f_X(x') + f_X(-x')) d\lambda(x').$$

If $x \leq 0$, then it is immediate that

$$P_{|X|}((-\infty, x]) = \mathbb{P}(|X| \leq x) = 0 = \int_{(-\infty,x]} 0 \, d\lambda.$$

We proved the equality of measures for all Borel sets of the form $(-\infty, x]$, so we proved it for all Borel sets. Indeed $P_{|X|} \in \mathcal{P}^+_{a.c}$.
The last statement of this theorem is a direct consequence of the definition of a mantissa, see Definition 3.1. The mantissa of a number $x \in \mathbb{R}$ is equal to the mantissa of $|x|$.

$\square$

Lemma 6.2 leads easily to the following lemma.

Lemma 6.4 Let $P_X \in \mathcal{P}^+$, then $P_{m_b(X)} \in \mathcal{P}^+$, and the mantissa distribution relative to base b is

$$P_{m_b(X)} = P_{b^{\log_b(X) \pmod 1}}.$$

Proof In the discussion just before Definition 5.1 we already saw that $P_{m_b(X)} \in \mathcal{P}$. We assumed X to be a strictly positive random variable, so $m_b(X)$ is also strictly positive. We conclude $P_{m_b(X)} \in \mathcal{P}^+$. To finish the proof, we prove the second statement of the lemma. So, we have to prove that for all Borel sets $B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(\{\omega \in \Omega : m_b(X(\omega)) \in B\}) = \mathbb{P}\Big(\{\omega \in \Omega : b^{\log_b(X(\omega)) \ (\mathrm{mod}\ 1)} \in B\}\Big).$$

Given is that $\mathbb{P}(\{\omega \in \Omega : X(\omega) \leq 0\}) = 0$, namely $P_X \in \mathcal{P}^+$. Now

$$\mathbb{P}(\{\omega \in \Omega : m_b(X(\omega)) \in B\}) =$$

$$\mathbb{P}(\{\omega \in \Omega : m_b(X(\omega)) \in B \ \text{ and } \ X(\omega) > 0\}) + \mathbb{P}(\{\omega \in \Omega : m_b(X(\omega)) \in B \ \text{ and } \ X(\omega) \leq 0\}) =$$

$$\mathbb{P}\Big(\{\omega \in \Omega : b^{\log_b(X(\omega)) \ (\mathrm{mod}\ 1)} \in B\}\Big) + 0.$$

□

The next theorem is the main theorem of this section.

Theorem 6.5 (Main Theorem) Let $P_X \in \mathcal{P}^+$. Let $b > 1$ be a real base. Then the mantissa distribution function $F_{m_b(X)}$ can be expressed in terms of the distribution function F_X:

$$F_{m_b(X)}(x) = \begin{cases} 0 & \text{for } x < 1, \\ \sum_{k=-\infty}^{\infty} \Big(F_X(x \cdot b^k) - F_X^-(b^k)\Big) & \text{for } x \in [1,b) \text{ and} \\ 1 & \text{for } x \geq b. \end{cases}$$

Let $P_X \in \mathcal{P}^+_{a.c.}$ with density function f_X. Then $P_{m_b(X)} \in \mathcal{P}^+_{a.c.}$ and the mantissa density function $f_{m_b(X)}$ can be expressed in terms of f_X:

$$f_{m_b(X)}(x) = \begin{cases} \sum_{k=-\infty}^{\infty} b^k \cdot f_X(b^k \cdot x) & \text{for a.e. } x \in [1,b) \ \text{ and} \\ 0 & \text{for a.e. other } x. \end{cases}$$

Remark that $F_X^-(b^k)$ is the left limit of F_X in the point b^k. In case F_X is left-continuous in b^k, we surely have $F_X^-(b^k) = F_X(b^k)$.

In order to prove this theorem, we need some lemmata. The proof is based on Lemma 6.4. Given P_X we will successively look at $P_{\log_b(X)}$, $P_{\log_b(X)(\mathrm{mod}\ 1)}$ and $P_{b^{\log_b(X)(\mathrm{mod}\ 1)}} = P_{m_b(X)}$.

Lemma 6.6 Let $P_X \in \mathcal{P}^+$. Let $b > 1$ be a real base. Then $P_{\log_b(X)} \in \mathcal{P}$ and its distribution function can be expressed in terms of the distribution function F_X:

$$F_{\log_b(X)}(x) = F_X(b^x) \qquad \text{for } x \in \mathbb{R}.$$

Let $P_X \in \mathcal{P}^+_{a.c.}$ with density function f_X. Then $P_{\log_b(X)} \in \mathcal{P}_{a.c.}$ and its density function can be expressed in terms of f_X:

$$f_{\log_b(X)}(x) = \log_e(b) \cdot b^x \cdot f_X(b^x) \qquad \text{for almost every } x \in \mathbb{R}.$$

Proof From $P_X \in \mathcal{P}^+$ we know that X is positive. The function $\log_b(x)$ is a Borel function, so $\log_b(X)$ is a random variable. Theorem 4.3 implies that $P_{\log_b(X)}$ is a probability measure. So, $P_{\log_b(X)} \in \mathcal{P}$. For the distribution function we can write

$$F_{\log_b(X)}(x) = P_{\log_b(X)}((-\infty,x])) = P_X((0,b^x]) = F_X(b^x) - F_X(0) = F_X(b^x).$$

Now assume that $P_X \in \mathcal{P}^+_{a.c.}$ and let B be a Borel set,

$$P_{\log_b(X)}(B) = P_X(\{x \in \mathbb{R} : \log_b(x) \in B\}) = \int_{\{x \in \mathbb{R}: \log_b(x) \in B\}} f_X(x) d\lambda = \int_B f_X(b^x) \cdot \log_e(b) \cdot b^x d\lambda,$$

where the last step is obtained by a change of variables. So, indeed $P_{\log_b(X)} \in \mathcal{P}_{a.c.}$.

$\square$

In Lemma 6.8 we will encounter an infinite sum and an integral which we would like to interchange. Therefore we will use the Monotone Convergence Theorem (see [WIL]), which we will state here.

Theorem 6.7 (Monotone-Convergence Theorem) Let (S,Σ,μ) be a measure space. If (f_n) is a sequence of non-negative Σ-measurable functions such that $f_n \uparrow f$, then

$$\int_S f_n(s)\mu(ds) \uparrow \int_S f(s)\mu(ds).$$

Now we are ready to give the next lemma.

Lemma 6.8 Let $b > 1$ be a real base. Let $P_{\log_b(X)} \in \mathcal{P}$. Then $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}$ and its distribution function can be expressed in terms of the distribution function $F_{\log_b(X)}$:

$$F_{\log_b(X)(\text{mod } 1)}(x) = \begin{cases} 0 & \text{for } x<0, \\ \sum_{k=-\infty}^{\infty}(F_{\log_b(X)}(x+k) - F^-_{\log_b(X)}(k)) & \text{for } x\in[0,1), \\ 1 & \text{for } x\geq 1. \end{cases}$$

Let $P_{\log_b(X)} \in \mathcal{P}_{a.c.}$. Then $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}_{a.c.}$ and its density function can be expressed in terms of the density function $f_{\log_b(X)}$,

$$f_{\log_b(X)(\text{mod } 1)}(x) = \begin{cases} \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k) & \text{for a.e. } x\in[0,1), \\ 0 & \text{for a.e. other } x. \end{cases}$$

Proof Since $P_{\log_b(X)} \in \mathcal{P}$, we know that $\log_b(X)$ is a random variable. This implies that $\log_b(X)(\text{mod } 1)$ is a random variable. Theorem 4.3 implies that $P_{\log_b(X)(\text{mod } 1)}$ is a probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and we write $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}$. The distribution function is

$$F_{\log_b(X)(\text{mod } 1)}(x) = P_{\log_b(X)(\text{mod } 1)}((-\infty,x])) = \begin{cases} P_{\log_b(X)}(\emptyset) & \text{if } x<0, \\ P_{\log_b(X)}(\bigcup_{k=-\infty}^{\infty}[k,k+x]) & \text{if } x\in[0,1), \\ P_{\log_b(X)}(\mathbb{R}) & \text{if } x\geq 1. \end{cases}$$

Of course we have $P_{\log_b(X)}(\emptyset) = 0$ and $P_{\log_b(X)}(\mathbb{R}) = 1$. In the second case we have

$$P_{\log_b(X)}\Big(\bigcup_{k=-\infty}^{\infty}[k,k+x]\Big) = \sum_{k=-\infty}^{\infty} P_{\log_b(X)}\Big([k,k+x]\Big) = \sum_{k=-\infty}^{\infty} P_{\log_b(X)}\Big((-\infty,k+x]\backslash(-\infty,k)\Big)$$

$P_{\log_b(X)}$ is a finite measure, so this is equal to

$$\sum_{k=-\infty}^{\infty}\Big(P_{\log_b(X)}((-\infty,k+x]) - P_{\log_b(X)}((-\infty,k))\Big) = \sum_{k=-\infty}^{\infty}\Big(F_{\log_b(X)}(k+x) - F^-_{\log_b(X)}(k)\Big)$$

We have proved the formula for distribution functions. Let $P_{\log_b(X)} \in \mathcal{P}_{a.c.}$. Let $B\in\mathcal{B}$, then

$$P_{\log_b(X)(\text{mod } 1)}(B) = P_{\log_b(X)}\Big(\bigcup_{k=-\infty}^{\infty}(B\cap[0+k,1+k))\Big) = \int_{\bigcup_{k=-\infty}^{\infty}(B\cap[0+k,1+k))} f_{\log_b(X)}(x)d\lambda(x) =$$

$$\sum_{k=-\infty}^{\infty}\int_{(B\cap[0+k,1+k))} f_{\log_b(X)}(x)d\lambda(x) = \sum_{k=-\infty}^{\infty}\int_{(B\cap[0,1))} f_{\log_b(X)}(x+k)d\lambda(x) =$$

$$\sum_{k=-\infty}^{\infty}\int 1_{B\cap[0,1)}\cdot f_{\log_b(X)}(x+k)d\lambda(x)$$

Now we would like to interchange sum and integral. Consider for every $n \geq 0$ the function

$$\text{sum}_n(x) = \sum_{k=-n}^{n} 1_{B\cap[0,1)} \cdot f_{\log_b(X)}(x+k).$$

This sequence of functions is growing to

$$\text{sum}(x) = \sum_{k=-\infty}^{\infty} 1_{B\cap[0,1)} \cdot f_{\log_b(X)}(x+k).$$

Now we will use the Monotone-Convergence Theorem (6.7). Let in the theorem $(S,\Sigma,\mu) = (\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$ and let $f_n = \text{sum}_n$. This functions are finite sums of Borel functions, so they are Borel functions. Let $f = \text{sum}$. Then according to the Monotone-Convergence Theorem, we have

$$\int_{\mathbb{R}} \text{sum}_n d\lambda \uparrow \int_{\mathbb{R}} \text{sum} d\lambda,$$

what implies that

$$\lim_{n\to\infty} \sum_{k=-n}^{n} \int_{\mathbb{R}} 1_{B\cap[0,1)} \cdot f_{\log_b(X)}(x+k)d\lambda(x) = \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{k=-n}^{n} 1_{B\cap[0,1)} \cdot f_{\log_b(X)}(x+k)d\lambda(x) =$$

$$\lim_{n\to\infty} \int_{\mathbb{R}} \text{sum}_n d\lambda = \int_{\mathbb{R}} \text{sum} d\lambda =$$

$$\int_{\mathbb{R}} \sum_{k=-\infty}^{\infty} 1_{B\cap[0,1)} \cdot f_{\log_b(X)}(x+k)d\lambda(x) = \int_B 1_{[0,1)} \cdot \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k)d\lambda(x).$$

We obtained:

$$P_{\log_b(X)(\text{mod } 1)}(B) = \int_B 1_{[0,1)} \cdot \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k)d\lambda(x) \qquad \text{for } B \in \mathcal{B}.$$

We proved $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}_{a.c.}$. Remark that the function $1_{[0,1)} \cdot \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k)$ may be infinity, so according to our definition of a density function in 4.3 (III) it may possibly not be a density function. But as long as $P_{\log_b(X)(\text{mod } 1)}(\mathbb{R}) = \int_{\mathbb{R}} 1_{[0,1)} \cdot \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k)d\lambda(x) = 1$, the set where it is infinity is a set with Lebesgue measure zero. We can define a density function of $\log_b(X)(\text{mod } 1)$ as follows

$$f_{\log_b(X)(\text{mod } 1)}(x) = \begin{cases} 1_{[0,1)} \cdot \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k) & \text{for all } x \in \mathbb{R} \text{ where the sum is finite} \\ 0 & \text{otherwise.} \end{cases}$$

□

Lemma 6.9 Let $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}$. Then $P_{m_b(X)} \in \mathcal{P}^+$ and its distribution function can be expressed in terms of the distribution function of $P_{\log_b(m_b(X))} = P_{\log_b(X)(\text{mod } 1)}$ as follows:

$$F_{m_b(X)}(x) = \begin{cases} 0 & \text{for } x < 1, \\ F_{\log_b(X)(\text{mod } 1)}(\log_b(x)) & \text{for } x \in [1,b), \\ 1 & \text{for } x \geq b. \end{cases}$$

Let $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}_{a.c.}$, then $P_{m_b(X)} \in \mathcal{P}^+_{a.c.}$ and its density function can be expressed in terms of the density function of $P_{\log_b(X)(\text{mod } 1)}$ as follows:

$$f_{m_b(X)}(x) = \begin{cases} \frac{f_{\log_b(X)(\text{mod } 1)}(\log_b(x))}{x \cdot \log_e(b)} & \text{for a.e } x \in [1,b), \\ 0 & \text{for a.e other } x. \end{cases}$$

Proof We already know that $P_{m_b(X)} \in \mathcal{P}$. Let us look at the distribution function of $P_{m_b(X)}$. For $x \geq 1$ we have

$$F_{m_b(X)}(x) = P_{m_b(X)}((-\infty,x]) = P_{\log_b(m_b(X))}((-\infty,\log_b(x)]) =$$
$$P_{\log_b(X)(\text{mod } 1)}((-\infty,\log_b(x)]) = F_{\log_b(X)(\text{mod } 1)}(\log_b(x)).$$

If $x \geq b$, then $\log_b(x) \geq 1$ and Lemma 6.8 implies that $F_{\log_b(X)(\text{mod } 1)}(\log_b(x)) = 1$. If $x < 1$, we have $F_{m_b(X)}(x) = P_{m_b(X)}((-\infty,x]) = 0$, because we know that $x' > 0 \Rightarrow m_b(x') \in [1,b)$. We have $P_{m_b(X)} \in \mathcal{P}^+$.

Concerning the density functions, we see that for a Borel set B, we have

$$P_{m_b(X)}(B) = P_{\log_b(m_b(X))}(\{x \in \mathbb{R} : b^x \in B\}) = P_{\log_b(X)(\text{mod } 1)}(\{x \in \mathbb{R} : b^x \in B\}) =$$
$$\int_{\{x\in\mathbb{R}:b^x\in B\}} f_{\log_b(X)(\text{mod } 1)}(x)d\lambda(x) = \int_B f_{\log_b(X)(\text{mod } 1)}(\log_b(x)) \cdot \frac{1}{x \cdot \log_e(b)} d\lambda(x),$$

where the last step is obtained by applying a change of variables. So, indeed $P_{\log_b(X)(\text{mod } 1)} \in \mathcal{P}_{a.c.} \Rightarrow P_{m_b(X)} \in \mathcal{P}_{a.c.}$. Further from $F_{m_b(X)}(x) = 0$ for $x < 1$ it follows that $f_{m_b(X)}(x) = 0$ for almost every $x < 1$and from $F_{m_b(X)}(x) = 1$ for $x \geq b$ it follows that $f_{m_b(X)}(x) = 0$ for almost every $x \geq b$. We have $P_{m_b(X)} \in \mathcal{P}^+_{a.c.}$.

□

Proof of Theorem 6.5 Let us look at the mantissa distribution function relative to base b. Let $x \in [1,b)$, then

$$F_{m_b(X)}(x) = F_{\log_b(X)(\text{mod } 1)}(\log_b(x)) = \sum_{k=-\infty}^{\infty} (F_{\log_b(X)}(\log_b(x)+k) - F^-_{\log_b(X)}(k)) =$$
$$\sum_{k=-\infty}^{\infty} (F_X(b^{\log_b(x)+k}) - F^-_X(b^k)) = \sum_{k=-\infty}^{\infty} (F_X(x \cdot b^k) - F^-_X(b^k)).$$

In Lemma 6.9 we already saw that $F_{m_b(X)}(x) = 0$ for $x < 1$ and $F_{m_b(X)}(x) = 1$ for $x \geq b$.

Now we will prove the statement about density functions. Let $x \in [1,b)$, we derive

$$f_{m_b(X)}(x) = \frac{f_{\log_b(X)(\text{mod } 1)}(\log_b(x))}{x \cdot \log_e(b)} = \frac{1}{x \cdot \log_e(b)} \cdot \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(\log_b(x)+k) =$$
$$\frac{1}{x \cdot \log_e(b)} \cdot \sum_{k=-\infty}^{\infty} \log_e(b) \cdot b^{\log_b(x)+k} \cdot f_X(b^{\log_b(x)+k}) = \frac{1}{x \cdot \log_e(b)} \cdot \sum_{k=-\infty}^{\infty} \log_e(b) \cdot x \cdot b^k \cdot f_X(x \cdot b^k) =$$
$$\sum_{k=-\infty}^{\infty} b^k \cdot f_X(x \cdot b^k) \qquad \text{almost surely.}$$

The first equality follows from Lemma 6.9. The second equality follows from Lemma 6.8. The third equality follows from Lemma 6.6. And the last two equalities follow by calculation. If $x < 1$ or $x \geq b$, then we know by Lemma 6.9 that $f_{m_b(X)}(x) = 0$ almost surely. Remark that for $x \in [1,b)$ the sum $\sum_{k=-\infty}^{\infty} b^k \cdot f_X(x \cdot b^k)$ can be infinity. Fortunately this happens if and only if for $y = \log_b(x)$ the sum $\sum_{k=-\infty}^{\infty} f_{\log_b(X)(\text{mod } 1)}(y+k)$ is infinity, which almost surely does not happen. So, we can define a density function of $P_{m_b(X)}$ by

$$f_{m_b(X)}(x) = \begin{cases} \sum_{k=-\infty}^{\infty} b^k \cdot f_X(b^k \cdot x) & \text{for } x \in [1,b) \text{ where the sum is finite,} \\ 0 & \text{otherwise.} \end{cases}$$

Indeed $P_{m_b(X)} \in \mathcal{P}^+_{a.c.}$.

□

6.1 THE STANDARD NORMAL DISTRIBUTION

Example 6.10 Let P_X be **the standard normal distribution** on $\mathbb{R}$. We know that $P_X \in \mathcal{P}_{a.c.}$. We are interested in the mantissa distribution of P_X relative to base 10. As long as the standard normal distribution is not in $\mathcal{P}^+_{a.c.}$ we have to change it such that it becomes in $\mathcal{P}^+_{a.c.}$. Since P_X is absolutely continuous with respect to Lebesgue measure, $P_X(\{0\}) = \mathbb{P}(X = 0) = 0$. So, without loss of generality we can assume that X is a random variable that is never zero. Then according to Theorem 6.3, $P_{|X|} \in \mathcal{P}^+_{a.c.}$ and by the same theorem the mantissa distribution of P_X is equal to the mantissa distribution of $P_{|X|}$. Now Lemma 6.4 tells us that

$$P_{m_{10}(X)} = P_{m_{10}(|X|)} = P_{10^{\log_{10}(|X|)(\text{mod }1)}}.$$

As long as P_X has a density function f_X, we can derive expressions for $f_{|X|}$, $f_{\log_{10}(|X|)}$, $f_{\log_{10}(|X|)(\text{mod }1)}$ and $f_{10^{\log_{10}(|X|)(\text{mod }1)}} = f_{m_{10}(|X|)} = f_{m_{10}(X)}$. We know that

$$f_X(x) = \frac{1}{\sqrt{2\cdot\pi}}e^{-x^2/2}, \qquad \text{for a.e. } x \in \mathbb{R},$$

so, by Theorem 6.3 we can express $f_{|X|}$ in terms of f_X as follows

$$f_{|X|}(x) = \frac{1}{\sqrt{2\cdot\pi}}e^{-x^2/2} + \frac{1}{\sqrt{2\cdot\pi}}e^{-(-x)^2/2} = 2\cdot\frac{1}{\sqrt{2\cdot\pi}}e^{-x^2/2}, \qquad \text{for a.e. } x > 0.$$

Look on the left side of Figure 1 for the graph of the density function of P_X and on the right side for the graph of the density function of $P_{|X|}$. By Lemma 6.6 and Lemma 6.8 we can express the density

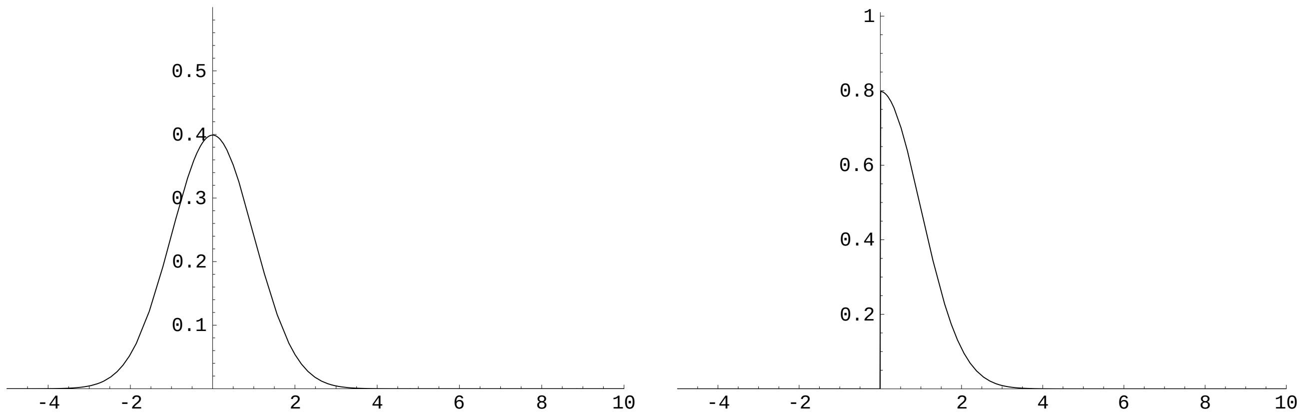


Figure 1: The standard normal distribution: On the left the graph of the density function of P_X and on the right the graph of the density function of $P_{|X|}$

functions $f_{\log_{10}(|X|)}$ and $f_{\log_{10}(|X|)(\text{mod }1)}$ as follows:

$$f_{\log_{10}(|X|)}(x) = \log_e(10)\cdot 10^x \cdot 2\cdot\frac{1}{\sqrt{2\cdot\pi}}e^{-(10^x)^2/2}, \qquad \text{for a.e. } x \in \mathbb{R}$$

and

$$f_{\log_{10}(|X|)(\text{mod }1)}(x) = \sum_{k=-\infty}^{\infty} \log_e(10)\cdot 10^x \cdot 2\cdot\frac{1}{\sqrt{2\cdot\pi}}e^{-(10^{x+k})^2/2}, \qquad \text{for a.e. } x \in [0,1)$$

Look on the left side of Figure 2 for the graph of the density function of $P_{\log_{10}(|X|)}$ and on the right side for the graph of the density function of $P_{\log_{10}(|X|)(\text{mod }1)}$.

In Figure 2 on the right the function is zero for $x < 0$ and $x \geq 1$, this is not plotted.

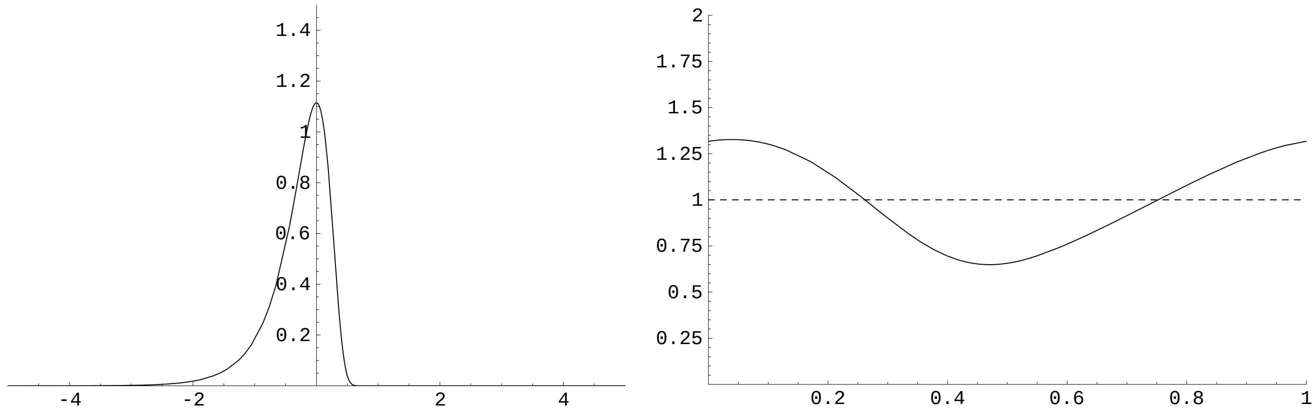


Figure 2: The standard normal distribution: On the left the graph of the density function of $P_{\log_{10}(|X|)}$ and on the right the graph of the density function of $P_{\log_{10}(|X|)(\text{mod } 1)}$.

According to the main theorem of this section, Theorem 6.5, we can express the density function of the mantissa distribution as follows:

$$f_{m_{10}(X)}(x) = f_{m_{10}(|X|)}(x) = \sum_{k=-\infty}^{\infty} 2 \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot 10^k \cdot e^{-(10^k \cdot x)^2/2}, \qquad \text{for a.e. } x \in [1, 10).$$

The computer helps us to draw the graph of this function, plotted in Figure 3.

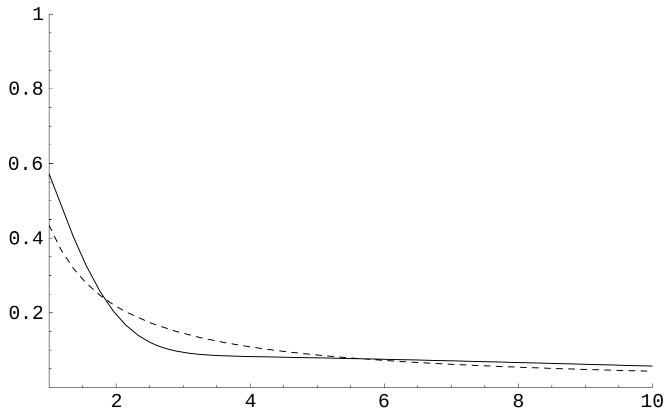


Figure 3: The standard normal distribution: The solid line is the graph of the mantissa density function relative to base 10 and the dotted line is the graph of $f_{BL_{10}}$

In the last image, Figure 3, the dotted line is the graph of the function $\frac{1}{x \cdot \log_e(10)}$. If the distribution satisfies Benford's Law in base 10, then the graph of $f_{m_{10}(X)}$ coincides almost everywhere with this line. We see that the mantissa density function of the standard normal distribution does not coincide almost everywhere with the dotted line and thus we conclude that the mantissae do not follow Benford's distribution for base 10. For numbers that follow the standard normal distribution we expect that $100 \cdot \int_1^2 f_{m_{10}(X)}(x) d\lambda(x) \approx 35.95$ percent has 1 as first significant digit. Let us calculate all the frequencies and put them next to Benford's frequencies in Table 2.

Table 2: The relative frequencies of the first digits of numbers that follow a the standard normal distribution compared with Benford's frequencies.

Digit		Benford	$N(0,1)$
1	$\cdots$	0.3010	0.3595
2	$\cdots$	0.1761	0.1290
3	$\cdots$	0.1249	0.0865
4	$\cdots$	0.0969	0.0810
5	$\cdots$	0.0792	0.0774
6	$\cdots$	0.0669	0.0734
7	$\cdots$	0.0580	0.0691
8	$\cdots$	0.0512	0.0644
9	$\cdots$	0.0458	0.0596

6.2 THE BETA DISTRIBUTION

Example 6.11 Let P_X be a **beta distribution**. We may assume that the random variable X is strictly positive such that $P_X \in \mathcal{P}^+_{a.c.}$. A density function of P_X is

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot (1-x)^{(\beta-1)} \cdot x^{(\alpha-1)} \cdot 1_{(0,1)}$$

By doing computer experiments one finds that for small α combined with large β this distribution has a mantissa distribution that is close to f_{BL_b}, where $b > 1$. We have to remark that the bigger the base, the smaller one has to choose α to get close to f_{BL_b}. Let us choose $\alpha = 0.1$ and $\beta = 4$. And let us plot the mantissa density function in base 10. Here follows, in Figure 4, the graph of this density function f_X, which is zero outside the interval $(0,1)$.

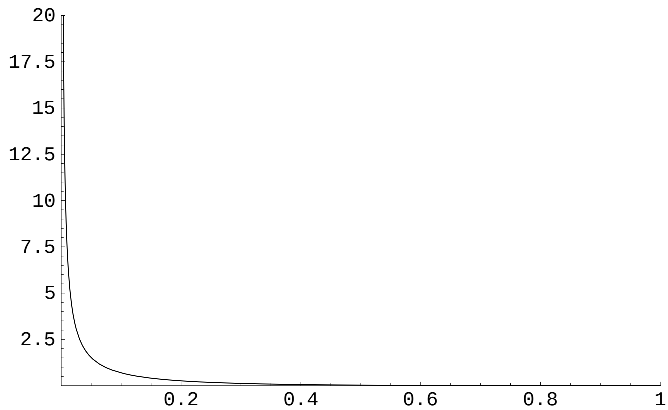


Figure 4: The beta distribution with $\alpha = 0.1$ and $\beta = 4$: The graph of the density function of P_X.

Look in Figure 5 on the left for the graph of the density function $f_{\log_{10}(X)}$ and on the right for the graph of the density function $f_{\log_{10}(X)(\mathrm{mod}1)}$. In figure 6 one can see the graph of the mantissa density function $f_{m_{10}(X)}$. The graph of $f_{m_{10}(X)}$ almost coincides with the graph of the function $\frac{1}{x\cdot\log_e(10)}$. We conclude that this beta distribution nearly satisfies Benford's Law. Mantissae of numbers following this beta distribution will nearly have relative frequencies as Benford's Law prescribes. Let us calculate all the frequencies and put them in Table 3. We see that the frequencies are indeed very close to the frequencies corresponding to Benford's Law. At the end of Section 11 we will shortly discuss the Beta distribution.

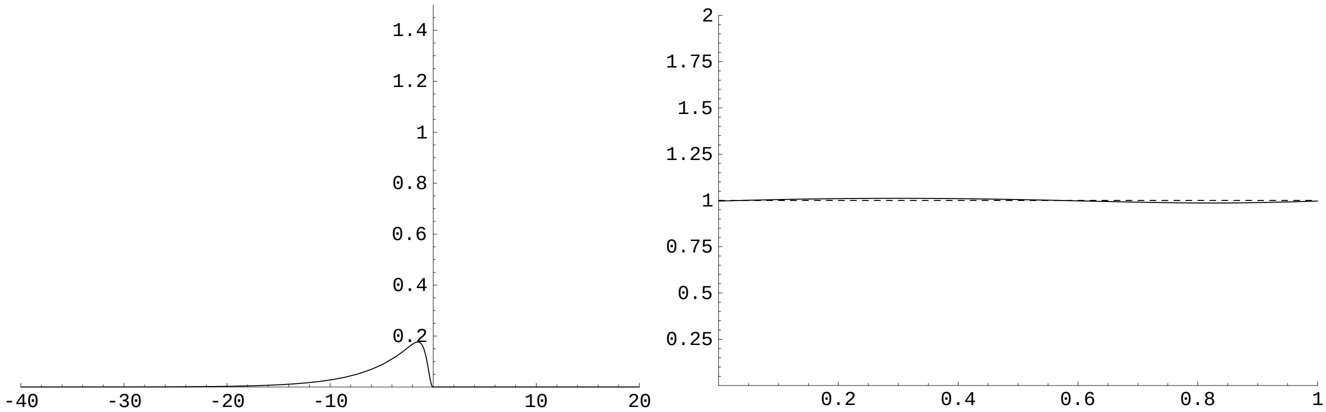


Figure 5: The beta distribution with $\alpha = 0.1$ and $\beta = 4$: On the left side the graph of the density function of $P_{\log_{10}(X)}$ and on the right side the graph of the density function of $P_{\log_{10}(X)(\mathrm{mod}1)}$.

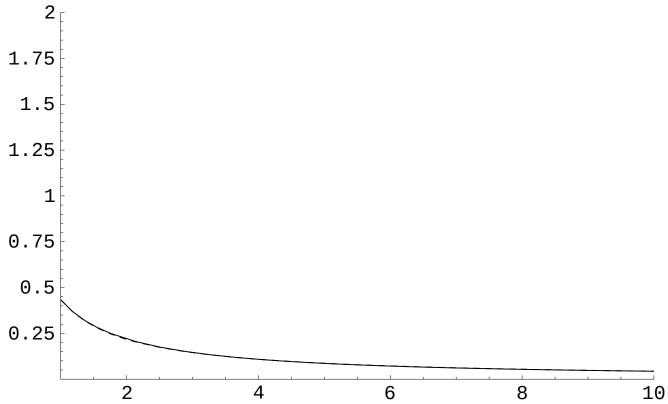


Figure 6: The beta distribution with $\alpha = 0.1$ and $\beta = 4$: The graph of the mantissa density function relative to base 10.

Table 3: The relative frequencies of the first digits of numbers that follow a beta distribution with $\alpha = 0.1$ and $\beta = 4$ compared with Benford's frequencies.

Digit		Benford	Beta$(0.1,4)$
1	⋯	0.3010	0.3030
2	⋯	0.1761	0.1778
3	⋯	0.1249	0.1252
4	⋯	0.0969	0.0963
5	⋯	0.0792	0.0783
6	⋯	0.0669	0.0660
7	⋯	0.0580	0.0572
8	⋯	0.0512	0.0507
9	⋯	0.0458	0.0455

7 Scalar multiplication

Raimi ([RAI] page 529) states:

> If the first digits of all the tables in the universe obey some fixed distribution law, Stigler's or Benford's or some other, that law must surely be independent of the system of units chosen, since God is not known to favor either the metric system or the English system. In other words, a universal first digit law, if it exists, must be scale-invariant.

This is a good motivation to study scale-invariance. Let $P_X \in \mathcal{P}^+_{a.c.}$. In this section we will examine the distribution of P_{cX}, where $c \in \mathbb{R}^+\backslash\{0\}$ is a scalar and X a random variable. It will be much easier to examine the distribution $P_{\log_b(m_b(cX))}$ or equivalently $P_{\log_b(cX)(\text{mod } 1)}$ instead of the distribution $P_{m_b(cX)}$.

In the next theorem we will see that the distribution of $P_{\log_b(m_b(X))} = P_{\log_b(X)(\text{mod } 1)}$ is uniform if and only if the distribution satisfies Benford's Law.

Theorem 7.1 (Uniform distribution on the circle) Let $b > 1$ and let $P_X \in \mathcal{P}^+_{a.c.}$. Then

$$f_{m_b(X)}(x) = \frac{1}{x \cdot \log_e(b)}, \text{for a.e. } x \in [1,b) \iff f_{\log_b(m_b(X))}(x) = 1, \text{ for a.e. } x \in [0,1).$$

Proof According to Lemma 6.9 we have for almost every $x \in [1,b)$ that

$$f_{m_b(X)}(x) = \frac{f_{\log_b(X) \ (\text{mod } 1)}(\log_b(x))}{x \cdot \log_e(b)}$$

Consider the next equivalencies:

$$f_{m_b(X)}(x) = \frac{1}{x \cdot \log_e(b)} \text{for a.e. } x \in [1,b) \iff \frac{f_{\log_b(X) \ (\text{mod } 1)}(\log_b(x))}{x \cdot \log_e(b)} = \frac{1}{x \cdot \log_e(b)} \text{for a.e. } x \in [1,b) \iff$$

$$f_{\log_b(X) \ (\text{mod } 1)}(\log_b(x)) = 1 \text{ for a.e. } x \in [1,b) \iff f_{\log_b(X) \ (\text{mod } 1)}(x) = 1 \text{ for a.e. } x \in [0,1)$$

□

Question: Why is this last theorem called *Uniform distribution on the circle*? Hint: what is the title of Section 6?

Let us examine the consequences of a multiplication with a scalar of the numbers of which we calculate the mantissae. The property that the logarithm of a product is a sum of logarithms will give us the solution to this problem. Let $c \in \mathrm{R}^+\backslash\{0\}$ and let f_X be the density function of P_X, then one easily checks that the density function of P_{cX} is equal to $\frac{1}{c} \cdot f(\frac{1}{c} \cdot x)$. The next theorem will show the consequences of a scalar multiplication for the density function of a logarithm of the mantissae.

Theorem 7.2 (Scalar multiplication) Let $b > 1$ and let $P_X \in \mathcal{P}^+_{a.c.}$ with density function f_X. Let $c \in \mathbb{R}^+\backslash\{0\}$. Now,

1. $f_{\log_b(m_b(cX))}(x) = f_{\log_b(m_b(X))}\Big((x - \log_b(c)) \ (\text{mod } 1)\Big)$ for almost every $x \in [0,1)$ and
2. $f_{\log_b(cX)}(x) = f_{\log_b(X)}(x - \log_b(c))$ for almost every $x \in \mathbb{R}$.

Proof First we prove the second statement. According to Lemma 6.6 we have

$$f_{\log_b(cX)}(x) = \log_e(b) \cdot b^x \cdot f_{cX}(b^x) \qquad \text{for a.e. } x \in \mathbb{R}.$$

The last expression is equal to

$$\log_e(b) \cdot b^x \cdot \frac{1}{c} \cdot f_X(\frac{1}{c} \cdot b^x) = \log_e(b) \cdot b^{(x - \log_b(c))} \cdot f_X(b^{(x - \log_b(c))}),$$

which is again by Lemma 6.6 almost surely equal to

$$f_{\log_b(X)}(x - \log_b(c)).$$

Now the first statement can be deduced as follows. Let $x \in [0,1)$ and let $a = (x - \log_b(c)) \pmod 1$ then $x - \log_b(c) = a + l$ with $a \in [0,1)$ and $l \in \mathbb{Z}$. Lemma 6.2 implies that $\log_b(m_b(cX)) = \log_b(cX) (\text{mod } 1)$. So, we have for almost every $x \in [0,1)$:

$$f_{\log_b(m_b(cX))}(x) = f_{\log_b(cX)(\text{mod } 1)}(x).$$

This is by Lemma 6.8 for almost every $x \in [0,1)$ equal to

$$\sum_{k=-\infty}^{\infty} f_{\log_b(cX)}(x+k).$$

This is by the second statement of this theorem for almost every $x \in [0,1)$ equal to

$$\sum_{k=-\infty}^{\infty} f_{\log_b(X)}((x+k) - \log_b(c)) = \sum_{k=-\infty}^{\infty} f_{\log_b(X)}((x - \log_b(c)) + k) =$$

$$\sum_{k=-\infty}^{\infty} f_{\log_b(X)}((a+l)+k) = \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(a + (l+k))$$

$$\sum_{j=-\infty}^{\infty} f_{\log_b(X)}(a+j).$$

The last step is true, because only the order in which the terms are added has been changed. The terms are all positive, so the order in which the terms are added does not matter. Which means that the sums are equal. The last expression is according to Lemma 6.8 for almost every $a \in [0,1)$ equal to

$$f_{\log_b(X)\ (\text{mod } 1)}(a) = f_{\log_b(X)\ (\text{mod } 1)}((x - \log_b(c))(\text{mod } 1)) = f_{\log_b(m_b(X))}((x - \log_b(c))(\text{mod } 1)).$$

Where the last equality is by Lemma 6.2 true for almost every $x \in [0,1)$.

$\square$

Example 7.3 In Example 6.10 we have determined the mantissa density function relative to base 10 of the standard normal distribution. Let P_X again be the standard normal distribution, where X is a random variable that is never zero. Here, in Figure 7, we plot again on the left the density function $f_{\log_{10}(m_{10}(X))}$ of $P_{m_{10}(X)}$, and on the right the mantissa density function relative to base 10, $f_{m_{10}(X)}$.
In Figure 8 we can see the consequence for $f_{\log_{10}(m_{10}(X))}$ and $f_{m_{10}(X)}$ of a scalar multiplication of $c =$

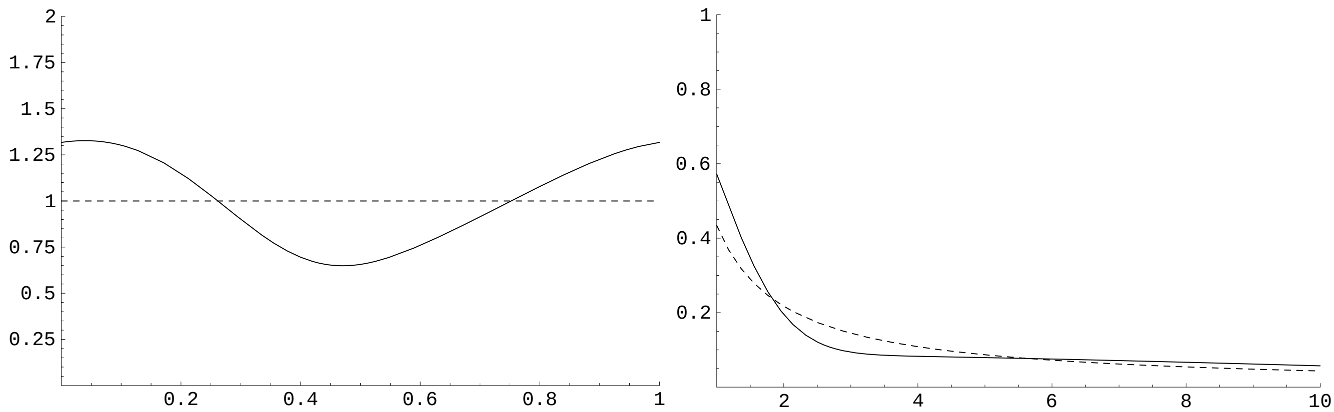


Figure 7: The standard normal distribution: On the left the graph of the density function of $P_{\log_{10}(X)\ (\text{mod } 1)}$ and on the right the graph of the density function of $P_{m_{10}}(X)$ (solid line) and the graph of $f_{BL_{10}}$ (dotted line).

$10^{(1/3)} \approx 2.2$ on X. According to Theorem 7.2 the graph of $f_{\log_{10}(m_{10}(X))}$ will rotate one third to the

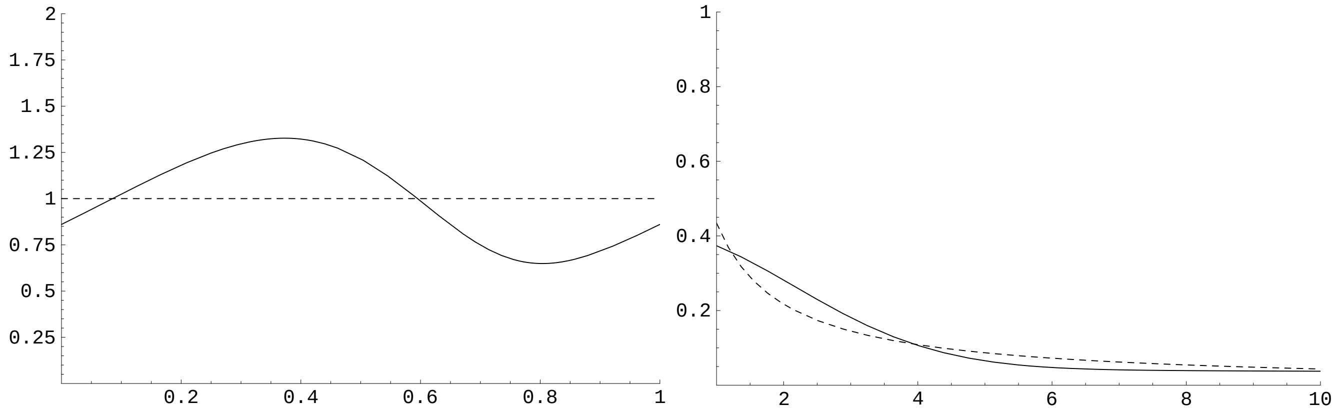


Figure 8: The standard normal distribution: On the left the graph of the density function $f_{\log_{10}(m_{10}(cX))}$ and on the right the graph of the density function $f_{m_{10}(cX)}$, with $c = 10^{(1/3)}$, (solid line) and the graph of $f_{BL_{10}}$ (dotted line).

right, while calculating modulo 1. Let P_{cX} be the distribution which has corresponding random variable $cX = 10^{(1/3)} \cdot X$. On the left we see the graph of the density function $f_{\log_{10}(m_{10}(cX))}$ and on the right the graph of the density function $f_{m_{10}(cX)}$. Next we will show the consequences of a scalar multiplication of $\tilde{c} = 10^{(2/3)} \approx 4.6$. For $f_{\log_{10}(m_{10}(X))}$ this yields a rotation of two third to the right. Let $P_{\tilde{c}X}$ be the distribution determined by $\tilde{c}X = 10^{(2/3)} \cdot X$. In Figure 9 we see on the left the graph of the density function $f_{\log_{10}(m_{10}(\tilde{c}X))}$ and on the right the graph of the density function $f_{m_{10}(\tilde{c}X)}$

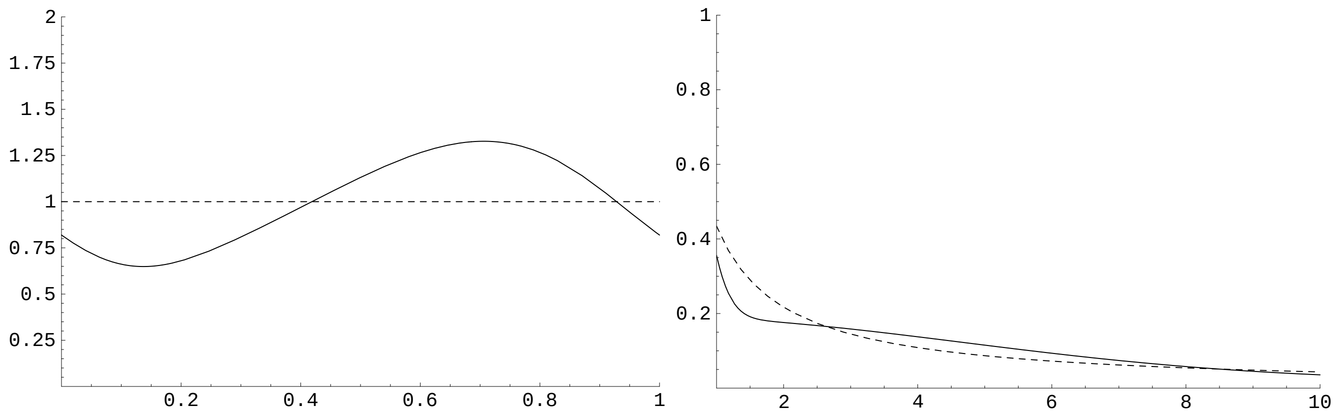


Figure 9: The standard normal distribution: On the left the graph of the density function $f_{\log_{10}(m_{10}(\tilde{c}X))}$ and on the right the graph of the density function $f_{m_{10}(\tilde{c}X)}$, with $\tilde{c} = 10^{(2/3)}$, (solid line) and the graph of $f_{BL_{10}}$ (dotted line).

Multiplication with $10^{3/3} = 10$ comes down to a rotation of 1 and so the mantissa density function will not change. This is evident, a multiplication of 10 comes down to shifting of the decimal point and will not change the mantissae. In the three figures on the left one can clearly see that the mean discrepancy of the dotted right line does not change, as a consequence of Theorem 8.6 of the next section this implies that also in the right figures the discrepancy between the mantissa density functions and the dotted line does not change. That means that the degree to which the mantissae are distributed logarithmically is invariant under scalar multiplication.

8 A BENFORD-DISTANCE ON DISTRIBUTIONS

To determine the degree to which a distribution satisfies Benford's Law, for integral bases one can calculate the frequencies of the significant digits and compare them with the frequencies given by Benford. It is also possible to define a distance on distributions which assigns to a distribution a value that indicates the degree in which it satisfies Benford's Law, this is possible for all real bases $b > 1$. Let us define the γ_b-distance of a distribution as follows.

Definition 8.1 Let $P_X \in \mathcal{P}^+_{a.c.}$ be a distribution with a mantissa density function $f_{m_b(X)}$. For every real base $b > 1$ the **γ_b-distance**, $\gamma_b : \mathcal{P}^+_{a.c.} \to [0, 2)$, is defined by

$$\gamma_b(P_X) = \int_{\mathbb{R}} |f_{m_b(X)} - f_{\mathrm{BL}_b}| d\lambda,$$

where f_{BL_b} is the density function corresponding to Benford's Law relative to base b, which is defined in Definition 5.2.

We see that the γ_b-distance is a $\mathcal{L}^1$-norm, we can write

$$\gamma_b(P_X) = \mathbb{E}(|f_{m_b(X)} - f_{\mathrm{BL}_b}|) = ||f_{m_b(X)} - f_{\mathrm{BL}_b}||.$$

The integral can be interpreted as the area between the mantissa density function of P_X and the density function of P_{BL_b}. It is a $\mathcal{L}^1$-norm, so the following theorem follows immediately.

Theorem 8.2 Let $P_X \in \mathcal{P}^+_{a.c.}$. Then P_X satisfies Benford's Law in base b if and only its γ_b-distance is 0.

If we say that a distribution is close to Benford's Law in a base b, we mean that the γ_b-distance is small. Using the triangle inequality one can show that the γ_b-distance of a combination of distributions is smaller or equal to the mean γ_b-distance of the separate distributions, see Section 8.3 on page 29. For all distributions P_X for which the γ_b-distance is defined holds $\gamma_b(P_X) \in [0, 2)$, see Theorem 8.11 on page 25.

Example 8.3 The normal distribution treated in Example 6.10 has γ_{10}-distance 0.2116. For the first significant digits this means that the total absolute difference between their frequencies and the frequencies in the table given by Newcomb, Table 1, lies between 0 and 0.2116, in fact the total difference is 0.2064. It can be the case that a distribution has a γ_{10}-distance not equal to zero, but has nevertheless the same percentages of first digits as in Newcomb's table. Of course, the first digits do not determine the mantissa distribution.

Example 8.4 The beta distribution treated in Example 6.11, where $\alpha = 0.1$ and $\beta = 4$, has γ_{10}-distance 0.0080.

The γ_b-distance can be calculated for all real bases bigger than 1. For positive integral bases it gives a bound for the maximum discrepancy between the frequencies of the significant digits and those given by Benford. In particular for base 10 we have the following theorem.

Theorem 8.5 Let $b = 10$ and let P_X be a distribution with mantissa density function $f_{m_{10}(X)}$, then the total absolute difference between the first significant digit frequencies of P_X and those given by Benford is bounded by the γ_{10}-distance of P_X. This means

$$\sum_{i=1}^{9} \left| \int_{[i,i+1]} f_{m_{10}(X)} d\lambda - \log_{10}(1 + i^{-1}) \right| \leq \gamma_{10}(P_X).$$

Proof

$$\sum_{i=1}^{9}\Big|\int_{[i,i+1]} f_{m_{10}(X)}d\lambda-\log_{10}(1+i^{-1})\Big|=\sum_{i=1}^{9}\Big|\int_{[i,i+1]} f_{m_{10}(X)}d\lambda-\int_{[i,i+1]} f_{\mathrm{BL}_{10}}d\lambda\Big|=$$

$$\sum_{i=1}^{9}\Big|\int_{[i,i+1]}(f_{m_{10}(X)}-f_{\mathrm{BL}_{10}})d\lambda\Big|\leq\sum_{i=1}^{9}\int_{[i,i+1]}\Big|f_{m_{10}(X)}-f_{\mathrm{BL}_{10}}\Big|d\lambda=$$

$$\int_{[1,10]}\Big|f_{m_{10}(X)}-f_{\mathrm{BL}_{10}}\Big|d\lambda=\gamma_{10}(P_X).$$

The inequality is a consequence of the known fact that if f is Lebesgue integrable, then $|\int_B f d\lambda|\leq\int_B|f|d\lambda$, with B a Borel set.

□

The next theorem states that the mean absolute discrepancy between $f_{\log_b(m_b(X))}$ and 1 is equal to the mean absolute discrepancy between $f_{m_b(X)}$ and f_{BL_b}.

Theorem 8.6 Let $P_X\in\mathcal{P}^+_{a.c.}$. We have

$$\gamma_b(P_X)=\int_{\mathbb{R}}|1_{[0,1)}-f_{\log_b(m_b(X))}|d\lambda$$

Proof Let $u=\log_b(v)$ then $v=b^u$ and u will run through the interval $[0,1)$ if v runs through the interval $[1,b)$. If we differentiate v with respect to u we get $\frac{dv}{du}=b^u\cdot\log_b(u)$. We have:

$$\gamma_b(P_X)=\int_{[1,b)}|\frac{1}{v\cdot\log_e(b)}-f_{m_b(X)}(v)|d\lambda(v).$$

Which is by the substitution rule for Lebesgue integrals equal to

$$\int_{[0,1)}|\frac{1}{b^u\cdot\log_e(b)}-f_{m_b(X)}(b^u)|\cdot b^u\cdot\log_e(b)d\lambda(u).$$

This is by Lemma 6.9 (and Lemma 6.2) equal to

$$\int_{[0,1)}|\frac{1}{b^u\cdot\log_e(b)}-\frac{f_{\log_b(m_b(X))}(\log_b(b^u))}{b^u\cdot\log_e(b)}|\cdot b^u\cdot\log_e(b)d\lambda(u).$$

We derive that this is equal to

$$\int_{[0,1)}|1-f_{\log_b(m_b(X))}(u)|d\lambda(u)=\int|1_{[0,1)}(u)-f_{\log_b(m_b(X))}(u)|d\lambda(u)$$

□

Remark that in general we see

$$\int_{\mathbb{R}}|F_{m_b(X)}-F_{\mathrm{BL}_b}|d\lambda\neq\int_{\mathbb{R}}|(x\cdot 1_{[0,1)}+1_{[1,\infty)})-F_{\log_b(m_b(X))}(x)|d\lambda(x).$$

We can use Theorem 8.6 to prove Theorem 8.8.

Definition 8.7 A function f : $\mathcal{P}^+_{a.c.}\to\mathbb{R}$ is called **scale-invariant** if

$$\mathrm{f}(P_{cX})=\mathrm{f}(P_X)$$

for all distributions $P_X\in\mathcal{P}^+_{a.c.}$ and all scalars $c\in\mathbb{R}^+\backslash\{0\}$.

Theorem 8.8 (Invariance under scalar multiplication) The γ_b-distance is scale-invariant, that is

$$\gamma_b(P_{cX}) = \gamma_b(P_X),$$

for all distributions $P_X \in \mathcal{P}^+_{a.c.}$ and all scalars $c \in \mathbb{R}^+ \backslash \{0\}$.

Proof By Theorem 8.6 we have

$$\gamma_b(P_{cX}) = \int_{\mathbb{R}} |1_{[0,1)} - f_{\log_b(m_b(cX))}(x)| d\lambda(x) = \int_{[0,1)} |1 - f_{\log_b(m_b(cX))}(x)| d\lambda(x).$$

This is by Theorem 7.2 equal to

$$\int_{[0,1)} \left| 1 - f_{\log_b(m_b(X))}((x - \log_b(c))(\text{mod } 1)) \right| d\lambda(x)$$

Let $a = \log_b(c)$. Then, for $x \in [0,1)$, $T_a(x) = (x-a)$ (mod 1) is λ-invariant . We can write the integral of the last display as

$$\int_{[0,1)} \left| 1 - f_{\log_b(m_b(X))} \right| \circ T_a(x) d\lambda(x)$$

The function $f_{\log_b(m_b(X))}$ is in $\mathcal{L}^1$ and so is $\left| 1 - f_{\log_b(m_b(X))} \right|$. Translation invariance mod 1 of Lebesgue measure on $[0,1)$, implies that the integral is equal to

$$\int_{[0,1)} |1 - f_{\log_b(m_b(X))}(x')| d\lambda(x') = \gamma_b(P_X).$$

$\square$

Definition 8.9 A distribution $P_X \in \mathcal{P}$ is called invariant under scalar multiplication if $P_{c \cdot X} = P_X$ for all $c > 0$.

Theorem 8.10 (Invariance under scalar multiplication $\iff$ Benford) Let $P_X \in \mathcal{P}^+_{a.c.}$. The mantissa distribution of P_X in base b is invariant under scalar multiplication if and only if P_X satisfies Benford's Law in base b

Proof "$\Rightarrow$" We have

$$P_{m_b(cX)} = P_{m_b(X)} \qquad \forall c > 0.$$

Then also

$$P_{\log_b(m_b(cX))} = P_{\log_b(m_b(X))} \qquad \forall c > 0.$$

We conclude that for all $c > 0$:

$$f_{\log_b(m_b(cX))}(x) = f_{\log_b(m_b(X))}(x) \qquad \text{for a.e. } x \in [0,1) \tag{1}$$

According to Theorem 7.2 we have for all $c > 0$:

$$f_{\log_b(m_b(cX))}(x) = f_{\log_b(m_b(X))}((x - \log_b(c)(\text{mod } 1)) = f_{\log_b(m_b(X))}(x) \qquad \text{for a.e. } x \in [0,1). \tag{2}$$

Now let $x, x' \in [0,1)$ be arbitrary. There is a $c \in \mathbb{R}^+ \backslash \{0\}$ such that $x' = (x - \log_b(c))$ (mod 1). Then with probability one

$$f_{\log_b(m_b(X))}(x') = f_{\log_b(m_b(cX))}(x) = f_{\log_b(m_b(X))}(x),$$

where the first equality follows from (2) and the second from (1). We conclude that $f_{\log_b(m_b(X))}$ is a.e. constant. For the density function $f_{\log_b(m_b(X))}$ we have $\int_{[0,1)} f_{\log_b(m_b(X))} d\lambda = 1$, so the constant must be 1. Now Theorem 7.1 implies that

$$f_{m_b(X)}(x) = \frac{1}{x \cdot \log_e(b)} \qquad \text{for a.e } x \in [1,b).$$

This means that

$$f_{m_b(X)}(x) = f_{\mathrm{BL}_b}(x) \qquad \text{for a.e. } x \in \mathbb{R}.$$

"$\Leftarrow$" We have

$$f_{m_b}(x) = \frac{1}{x \cdot \log_e(b)} \qquad \text{for a.e. } x \in [1,b).$$

Then according to Theorem 7.1 we have

$$f_{\log_b(m_b(X))}(x) = 1 \qquad \text{for a.e. } x \in [0,1).$$

Now $f_{\log_b(m_b(cX))}(x) = f_{\log_b(m_b(X))}(x - \log_b(c) \pmod 1) = 1 = f_{\log_b(m_b(X))}(x)$ for almost all $x \in [0,1)$. This means that

$$f_{m_b(cX)} = f_{m_b(X)} \qquad \text{a.e.}$$

Density functions determine probability distributions totally. So, this implies that

$$P_{m_b(cX)} = P_{m_b(X)} \qquad \text{for all } c > 0.$$

A scalar invariance holds.

□

Remark that Theorem 8.10 is always true, even without the assumption of absolute continuity, see [HI1].

Theorem 8.11 For all $P_X \in \mathcal{P}^+_{a.c.}$ the γ_b-distance is smaller than 2 and tends to 2 as b tends to. infinity.

Proof First we prove that

$$\lim_{b\to\infty} \gamma_b(P_X) = 2$$

Recall that according to Theorem 8.6 we have

$$\gamma_b(P_X) = \int_{\mathbb{R}} |1_{[0,1)} - f_{\log_b(m_b(X))}| d\lambda.$$

The distribution $P_{\log_b(m_b(X))}$ tends to the dirac measure δ_0 as b tends to infinity, for the following reason. For every $x \in (0,1)$: the measure that P_X assigns to $[\frac{1}{b^x}, b^x]$ tends to one as b tends to infinity. This means that the measure that $P_{\log_b(X)}$ assigns to $[-x,x]$ tends to 1 as b tends to infinity. This means that the measure that $P_{\log_b(m_b(X))}$ assigns to $[0,x] \cup [1-x,1)$ tends to 1 as b tends to infinity. Since we can make x as close to 0 as we want, we conclude that $P_{\log_b(m_b(X)}$ tends to the dirac measure δ_0.

We have for all $0 < x < 1/2$ that

$$\lim_{b\to\infty} P_{\log_b(m_b(X))}\Big([0,x] \cup [1-x,1)\Big) = 1.$$

Then we also have that for all $0 < x < 1/2$:

$$\lim_{b\to\infty} \int_{(x,1-x)} f_{\log_b(m_b(X))} d\lambda = 0 \tag{3}$$

and

$$\lim_{b\to\infty} \int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))} d\lambda = 1 \tag{4}$$

From (3) and the fact that $f_{\log_b(m_b(X))}$ is nonnegative we derive that for all $0 < x < 1/2$:

$$\lim_{b\to\infty} \int_{(x,1-x)} f_{\log_b(m_b(X))} \cdot 1_{\{f_{\log_b(m_b(X))} \le 1\}} d\lambda = 0 \tag{5}$$

and

$$\lim_{b\to\infty}\int_{(x,1-x)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda = 0 \tag{6}$$

and from (4) we derive

$$\lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda + \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}\leq 1\}}d\lambda = 1$$

For the right term we have

$$0 \leq \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}\leq 1\}}d\lambda \leq 2x \tag{7}$$

and conclude that

$$1-2x \leq \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda \leq 1. \tag{8}$$

From (5) and (6) we will derive that for all $0 < x < 1/2$:

$$\lim_{b\to\infty}\int_{(x,1-x)} |1-f_{\log_b(m_b(X))}|d\lambda = 1-2x \tag{9}$$

We derive it as follows:

$$\lim_{b\to\infty}\int_{(x,1-x)} |1-f_{\log_b(m_b(X))}|d\lambda = \lim_{b\to\infty}\int_{(x,1-x)} (1-f_{\log_b(m_b(X))})\cdot 1_{f_{\log_b(m_b(X))}\leq 1}d\lambda +$$

$$\lim_{b\to\infty}\int_{(x,1-x)} (f_{\log_b(m_b(X))}-1)\cdot 1_{f_{\log_b(m_b(X))}>1}d\lambda =$$

$$\lim_{b\to\infty}\int_{(x,1-x)} 1_{f_{\log_b(m_b(X))}\leq 1}d\lambda - \lim_{b\to\infty}\int_{(x,1-x)} f_{\log_b(m_b(X))}\cdot 1_{f_{\log_b(m_b(X))}\leq 1}d\lambda +$$

$$\lim_{b\to\infty}\int_{(x,1-x)} f_{\log_b(m_b(X))}\cdot 1_{f_{\log_b(m_b(X))}>1}d\lambda - \lim_{b\to\infty}\int_{(x,1-x)} 1_{f_{\log_b(m_b(X))}>1}d\lambda.$$

The second and third limits are both zero according to (5) and (6). And as long as $\lim_{b\to\infty}\lambda(\{f_{\log_b(m_b(X))} > 1\}) = 0$ and $\lim_{b\to\infty}\lambda(\{f_{\log_b(m_b(X))} \leq 1\}) = 1$, it follows that the first limit is equal to $1-2x$ and the fourth limit is equal to zero.

Using (8) we will see that

$$1-4x \leq \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} |1-f_{\log_b(m_b(X))}|d\lambda \leq 1+2x \tag{10}$$

First observe that

$$\lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} |1-f_{\log_b(m_b(X))}|d\lambda =$$

$$\lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} (1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}\leq 1\}}d\lambda + \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} (f_{\log_b(m_b(X))}-1)\cdot 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda$$

We have

$$0 \leq \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} (1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}\leq 1\}}d\lambda =$$

$$\lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} 1_{\{f_{\log_b(m_b(X))}\leq 1\}}d\lambda - \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}\leq 1\}}d\lambda \leq 2x$$

for the first limit and

$$\lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)}(f_{\log_b(m_b(X))}-1)\cdot 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda =$$

$$\lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} f_{\log_b(m_b(X))}\cdot 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda - \lim_{b\to\infty}\int_{[0,x]\cup[1-x,1)} 1_{\{f_{\log_b(m_b(X))}>1\}}d\lambda,$$

for the second limit, which according to (8) lies between $1-4x$ and 1.

Now (10) combined with (9) implies that

$$2-6x \leq \lim_{b\to\infty}\int_{[0,1]}|1-f_{\log_b(m_b(X))}|d\lambda \leq 2.$$

This holds for all $0<x<1/2$ and we conclude that $\lim_{b\to\infty}\gamma_b(P_X)=\lim_{b\to\infty}\int_{[0,1]}|1-f_{\log_b(m_b(X))}|d\lambda=2$.

The value 2 is also a bound for the γ_b-distance. For the density function $f_{\log_b(m_b(X))}$ holds

$$\int_{[0,1)} f_{\log_b(m_b(X))}d\lambda = 1.$$

From this follows that

$$\int_{[0,1)}(1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}<1\}}d\lambda = \int_{[0,1)}(f_{\log_b(m_b(X))}-1)\cdot 1_{\{f_{\log_b(m_b(X))}\geq 1\}}d\lambda.$$

We derive

$$\int_{[0,1)}|1-f_{\log_b(m_b(X))}|d\lambda = \int_{[0,1)}|(1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}<1\}}|d\lambda +$$

$$\int_{[0,1)}|(1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}\geq 1\}}|d\lambda = 2\cdot\int_{[0,1)}|(1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}<1\}}|d\lambda \leq 2.$$

The γ_b-distance cannot be 2, because then we would have $\int_{[0,1)}|(1-f_{\log_b(m_b(X))})\cdot 1_{\{f_{\log_b(m_b(X))}<1\}}|d\lambda = 1$, which means that $\lambda(\{x\in[0,1): f_{\log_b(m_b(X))}(x)=0\})=1$. This cannot be true because $f_{\log_b(m_b(X))}$ is a density function, which is 0 outside the interval $[0,1)$. We conclude $\gamma_b(P_X)<2$.

□

8.1 The normal distribution

Example 8.12 Let us examine the frequencies of digits of numbers that follow a **normal distribution** with mean μ and standard deviation σ. What happens for example if σ tends to infinity? Let P_X be a normal distribution with mean μ and standard deviation σ. Let f_X be a corresponding density function,

$$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-(x-\mu)^2/2\sigma^2}.$$

Here follows a theorem about the γ_b-degree to which a normal distribution satisfies Benford's Law.

Theorem 8.13 The γ_b-distance of a normal distribution is determined by the absolute value of the proportion between the mean and the standard deviation. For all $b>1$:

$$|\mu/\sigma| = |\mu'/\sigma'| \Longrightarrow \gamma_b(N(\mu,\sigma^2)) = \gamma_b(N(\mu',\sigma'^2)),$$

where $N(\mu,\sigma^2)$ and $N(\mu',\sigma'^2)$ are two normal distributions.

Proof Let $P_X = N(\mu, \sigma)$ and let us look at the distribution $P_{X/\sigma}$. Its density function is

$$f_{X/\sigma}(x) = \sigma \cdot f_X(\sigma \cdot x) = \frac{\sigma}{\sigma\sqrt{2\pi}} \cdot e^{-(\sigma \cdot x - \mu)^2/2\sigma^2} =$$

$$\frac{1}{\sqrt{2\pi}} \cdot e^{-(x-\mu/\sigma)^2/2}.$$

Then according to Theorem 8.8 we have $\gamma_b(P_{X/\sigma}) = \gamma_b(P_X)$. Theorem 6.3 implies that $P_{m_b(X)}$ is equal to $P_{m_b(|X|)}$, so only the absolute value of the quotient matters: $|\mu/\sigma|$.

□

This means for example that one can expect that the mantissae of numbers that follow a normal distribution with mean 100 and standard deviation 10 are distributed logarithmically in the same degree as the mantissae of numbers that follow a normal distribution with mean 1000 and standard deviation 100.
If we fix μ, then letting σ tend to infinity is the same as letting the quotient μ/σ tend to zero. So, the theorem has the following corollary.

Corollary 8.14

$$\lim_{\sigma \to \infty} \gamma_b(N(\mu, \sigma^2)) = \gamma_b(N(0,1)).$$

This means that for all μ the mantissa distribution of a normal distribution with σ tending to infinity gets as logarithmic as the mantissa distribution of a standard normal distribution, which is definitely not logarithmic.

In the next image, Figure 10, the γ-distance is plotted against the quotient μ/σ for base 10.
We can conclude that there does not exist a normal distribution with a logarithmic mantissa distribution. One can see easily that a normal distribution with quotient $\mu/\sigma = 0$ has no logarithmic distribution and the discrepancy increases if the quotient $|\mu/\sigma|$ increases. So, the standard normal distribution has the most logarithmic mantissa distribution of all normal distribution, accompanied by all other normal distributions with $\mu = 0$. We have the following conjecture:

Conjecture 8.15 With respect to the γ_b-distance, no normal distribution will satisfy Benford's Law better than the standard normal distribution $N(0,1)$. Moreover, for all real bases $b > 1$,

$$|\mu_1/\sigma_1| < |\mu_2/\sigma_2| \implies \gamma_b(N(\mu_1, \sigma_1)) < \gamma_b(N(\mu_2, \sigma_2))$$

8.2 The gamma distribution

Example 8.16 Let $P_X = \Gamma(\alpha, \lambda)$ be a gamma distribution and such that X is a strictly positive random variable. Then $P_X \in \mathcal{P}^+_{a.c.}$. It has parameters $\alpha > 0$ and $\lambda > 0$ and a density function f_X, given by

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\lambda x} \cdot 1_{x>0}.$$

Theorem 8.17 The scalar λ does not have influence on the γ_b-distance of the distribution.

Proof: Consider $P_{\lambda X}$. The density function of $P_{\lambda X}$ is $f_{\lambda X}(x) = \frac{1}{\lambda} \cdot f_X(\frac{1}{\lambda} \cdot x)$ or

$$f_{\lambda X}(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} \cdot e^{-x}.$$

Theorem 8.8 implies that $\gamma_b(P_{\lambda X}) = \gamma_b(P)$.

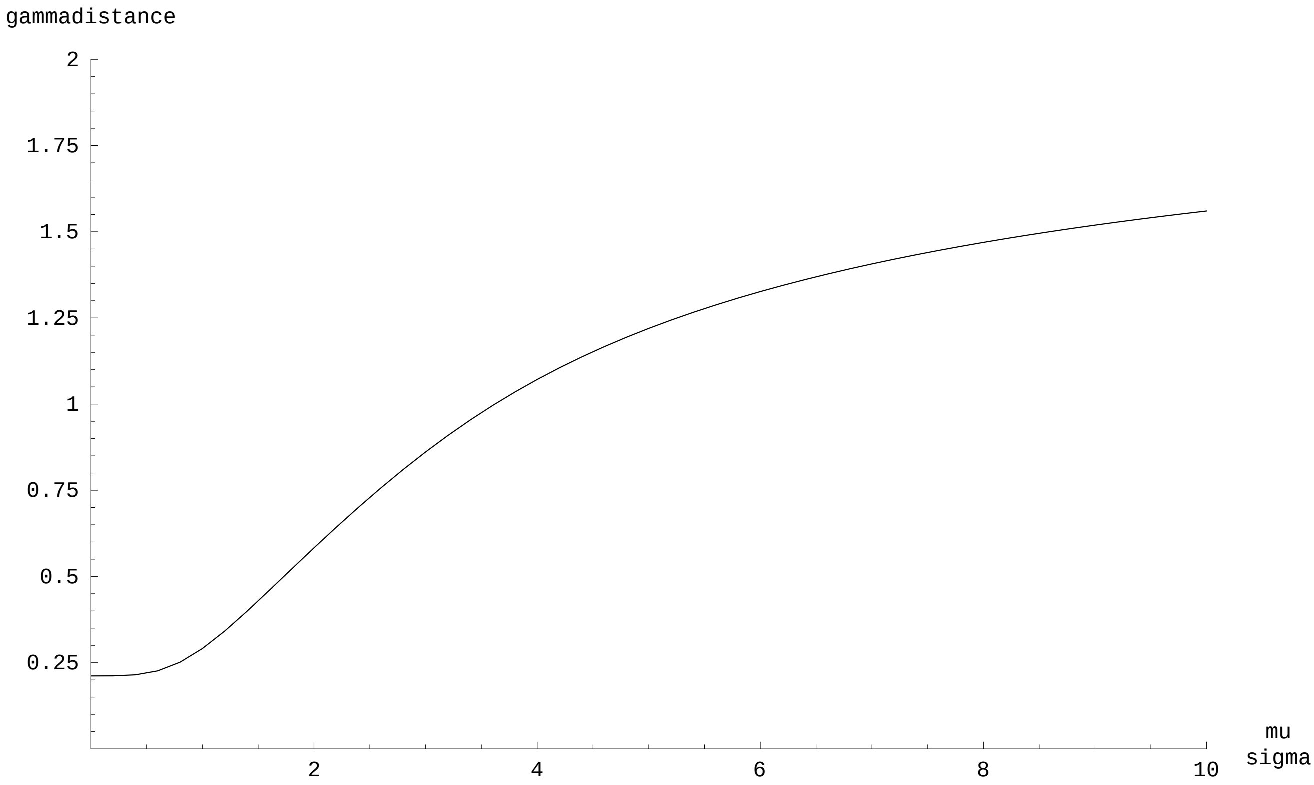


Figure 10: The γ_{10}-distance of a normal distribution with quotient μ/σ.

□

The exponential distribution is a gamma distribution with $\alpha = 1$. We have the following corollary of the theorem.

Corollary 8.18 For all $b > 1$, every **exponential distribution** has the same γ_b-distance, that is, does not depend on the mean of the distribution.

In Figure 11 the γ_{10}-distance of the gamma distribution is plotted against α.
What I cannot prove but did see while calculating mantissa distributions is that the γ_b-distance tends to 0 if α tends to 0, that it increases if α increases.

Conjecture 8.19 Let $\Gamma(\alpha,\lambda)$ be a a gamma distribution with parameters α and λ. For every b the γ_b-distance tends to zero if α tends to zero:

$$\lim_{\alpha \downarrow 0} \gamma_b(\Gamma(\alpha,\lambda)) = 0$$

and the γ_b-distance increases as α increases

$$\alpha_1 < \alpha_2 \Rightarrow \gamma_b(\Gamma(\alpha_1,\lambda))) < \gamma_b(\Gamma(\alpha_2,\lambda))).$$

8.3 Combinations of distributions

How could we explain that mantissae of numbers obtained from various sources, as in newspapers, often follow a logarithmic distribution. Ted Hill explained in his article A Statistical Derivation of the Significant-Digit Law [HI2] that the examining of numbers obtained from various sources can be simulated by looking at distributions that are a combination of various distributions. In newspapers numbers that follow a normal distribution stand next to numbers that follow uniform distributions, numbers that follow exponential distributions, etcetera.

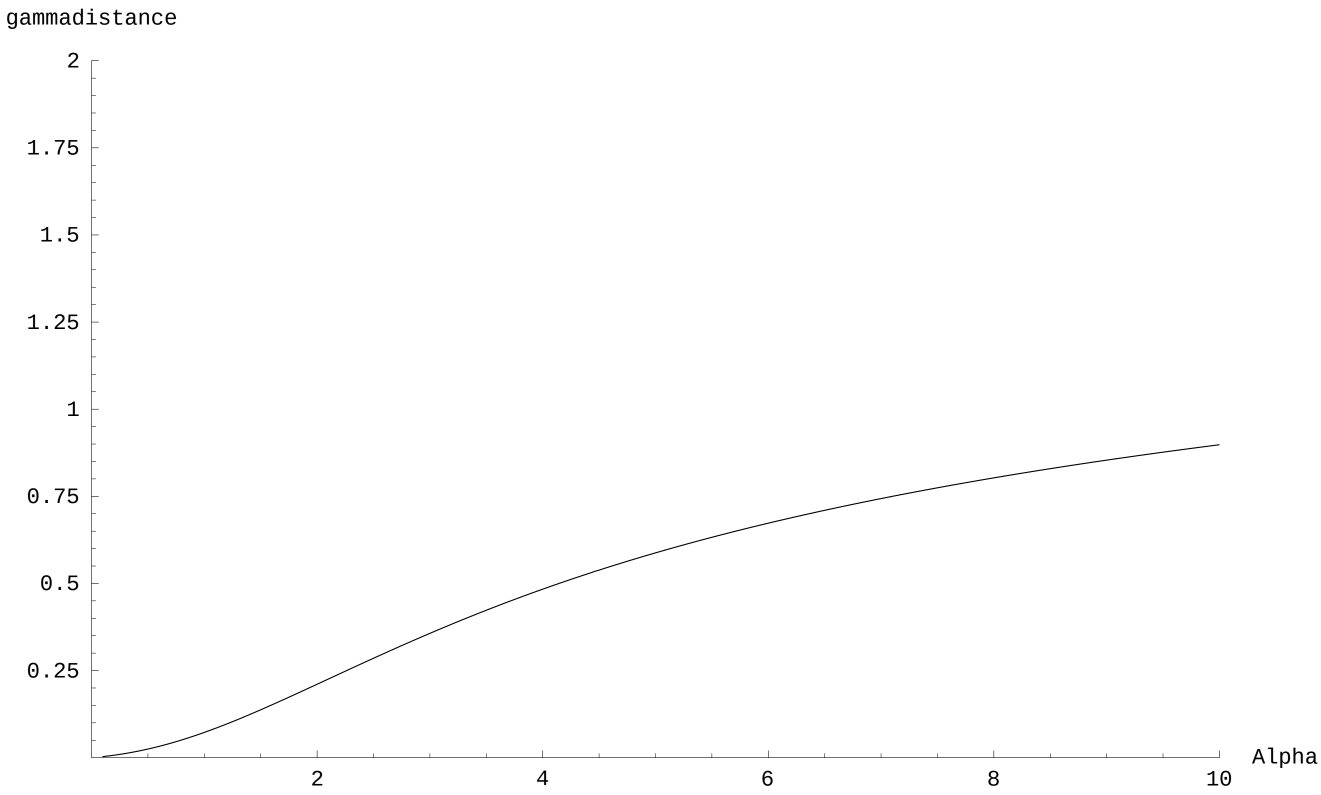


Figure 11: The γ_{10}-distance of a gamma distribution plotted against α.

Theorem 8.20 Let $P_X \in \mathcal{P}^+_{a.c.}$. Further, let $P_{X_k} \in \mathcal{P}^+_{a.c.}$ for $1 \leq k \leq n$ and such that

$$f_X = \frac{1}{n} \cdot \sum_{k=1}^{n} f_{X_k}.$$

Then, the γ_b-distance of the combination distribution is smaller or equal to the average γ-distance of the separate distributions:

$$\gamma_b(P_X) \leq \frac{1}{n} \cdot \sum_{k=1}^{n} \gamma_b(P_{X_k}).$$

Proof Given is that $f_X = \frac{1}{n} \cdot \sum_{k=1}^{n} f_{X_k}$. This implies that

$$f_{\log_b(X) \pmod 1} = \frac{1}{n} \cdot \sum_{k=1}^{n} f_{\log_b(X_k) \pmod 1}$$

Look at the probability space $\Big([0,1), \mathcal{B}([0,1)), \lambda\Big)$, and observe that $Y = 1 - f_{\log_b(X) \pmod 1}$ and $Y_k = 1 - f_{\log_b(X_k) \pmod 1}$ are random variables on this probability space (with expectation zero). According to Theorem 8.6 the γ-distances of P_X and P_{X_k} are $\mathbb{E}(|Y|)$ and $\mathbb{E}(|Y_k|)$, respectively. We use the triangle inequality to see that

$$\mathbb{E}(|Y_1 + Y_2 + \cdots + Y_n|) \leq \mathbb{E}(|Y_1|) + \mathbb{E}(|Y_2|) + \cdots + \mathbb{E}(|Y_n|).$$

From this follows that

$$\mathbb{E}(|Y|) = \mathbb{E}(|\frac{1}{n} \cdot \sum_{k=1}^{n} Y_k|) \leq \frac{1}{n} \cdot \sum_{k=1}^{n} \mathbb{E}(|Y_k|).$$

□

For an example concerning a combination of distributions see Example 11.5. Also compare Figure 22 with Figure 23.

9 Mantissa distribution in several bases

In 1981 P.Schatte wrote an article titled *On Random Variables with Logarithmic mantissa Distrution Relative to Several Bases*, see [SC1]. This article gives enough motivation to examine Benford's Law for several bases. The first thing I did in my research was examining the populations of 440 municipalities in The Netherlands. Of these 440 natural numbers, smallest 988 and biggest 739459 I calculated the mantissae for several bases and drew the empirical cumulative distribution functions (for a definition see for example page 346 of [RIC]). After that I calculated density functions of the logarithms of the mantissae, with respect to the same bases. I used a so called "kernel probability density estimate", where I used a standard normal function as weight function (for a definition see for example page 359 of [RIC]). Let us draw in figure 12 on the left the graphs of the **empirical cumulative distribution functions** of the mantissae and on the right the **density functions** of the logarithms of the mantissae with respect to base $b = 800000$, $b = 80000$, $b = 8000$, $b = 800$ and in figure 13 with respect to base $b = 80$, $b = 10$, $b = 8$, $b = 2$ and finally in figure 14 with respect to the real base $b = 1.8$. So, on the left we see an approximation of $F_{m_b(X)}$ and on the right an approximation of $f_{\log_b(m_b(X))}$.

Of course the mantissae of these numbers with respect to base $b = 800000$ are exactly the same as the numbers itself (first image in figure 12). In the first image of figure 12 on the right we see that the populations follow a log normal distribution, that means that the logarithms of the populations follow a normal distribution. For this base the discrepancy between the distribution of the mantissae and the logarithmic distribution is exactly the same as the discrepancy between the distribution of the numbers and the logarithmic distribution.
From the images on the left in figures 12, 13 and 14 we cannot get much information about the degree in which the mantissae are distributed logarithmically, but fortunately from the images on the right we can. We see that the top of the function moves to the right. This moving does not influence the degree. Interesting is that the top of the function decreases and the bell-shaped form stretches out, which makes the discrepancy between the graph of the function and the dotted straight line smaller, which means that the discrepancy between the distribution of the mantissae and the logarithmic distribution becomes smaller. If the base decreases, the two endpoints get together and the graph of the function starts to interfere with itself, what speeds up the process of convergence to the uniform distribution (see Figure 13). On the images on the right side for the bases smaller than 80 (see figure 13) the bell-shaped form is not recognizable, only if you knew it had that origin, you could see it. In the case $b = 10$ the distribution of the mantissae is close to logarithmic and in for $b = 2$ even more. For the small real base $b = 1.8$ we see in Figure 14 that the mantissae are distributed nearly logarithmically.
In this example we already saw that the distribution of the mantissae can dependent on the base in which you calculate in a very strong way. In this case the γ_b-distance appears to decrease as one lowers the base and tends to 0 if b tends to 1. A computer specialist who calculates in base 2 will claim that this numbers satisfy Benford's Law while a Maya-Indian who calculates in base 20 will claim the opposite: in this case they will never agree with each other. Before proving the convergence to the logarithmic distribution we need to develop some theory, sampling functions will appear to be a nice tool.

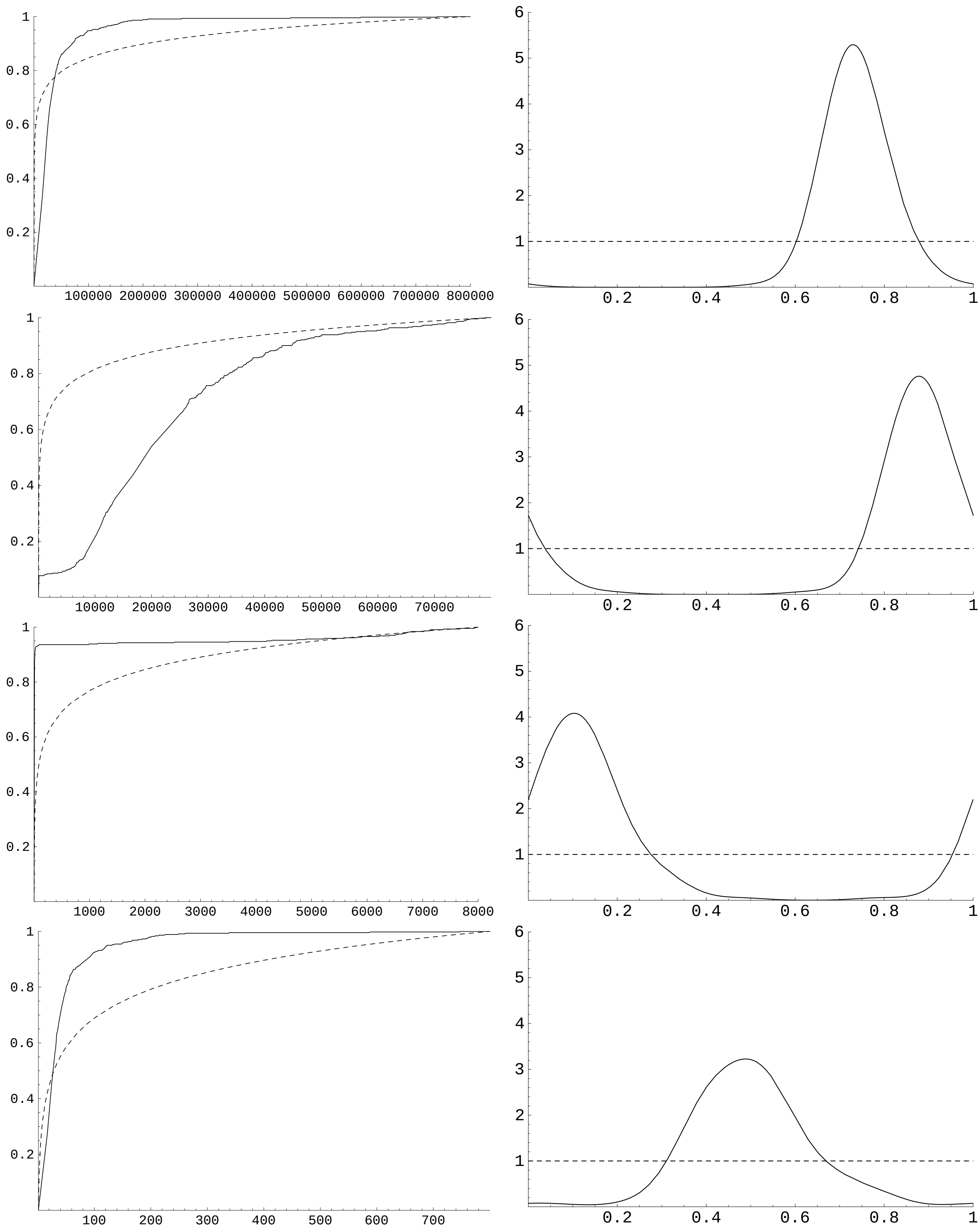


Figure 12: The Dutch municipalities: On the left the cumulative distribution functions of the mantissae (solid line) and F_{BL_b} (dotted line) and on the right the density functions of the logarithms of the mantissae relative to base 800000, 80000, 8000 and 800, respectively.

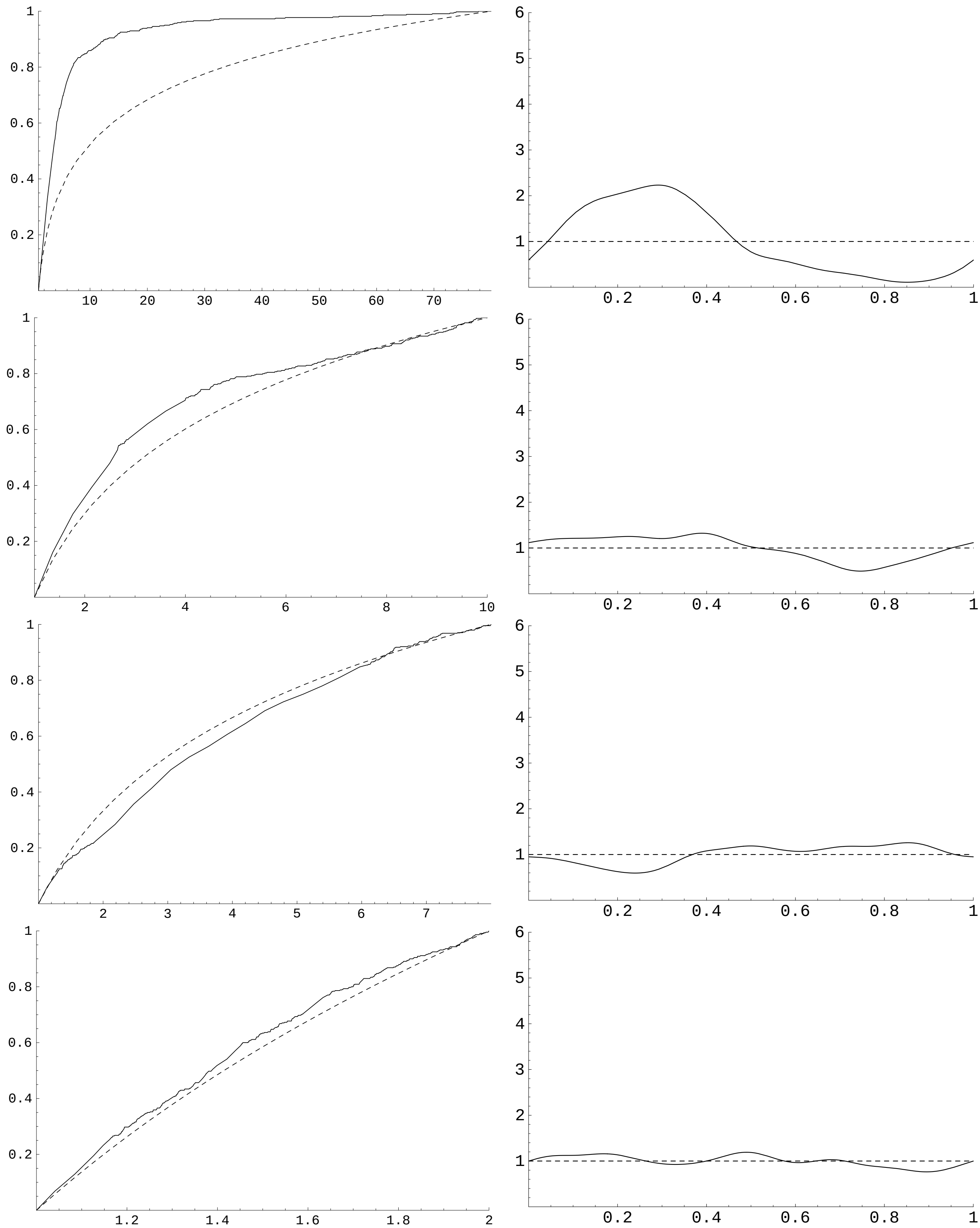


Figure 13: The Dutch municipalities: On the left the cumulative distribution functions of the mantissae (solid line) and F_{BL_b} (dotted line) and on the right the density functions of the logarithms of the mantissae relative to base 80, 10, 8 and 2, respectively.

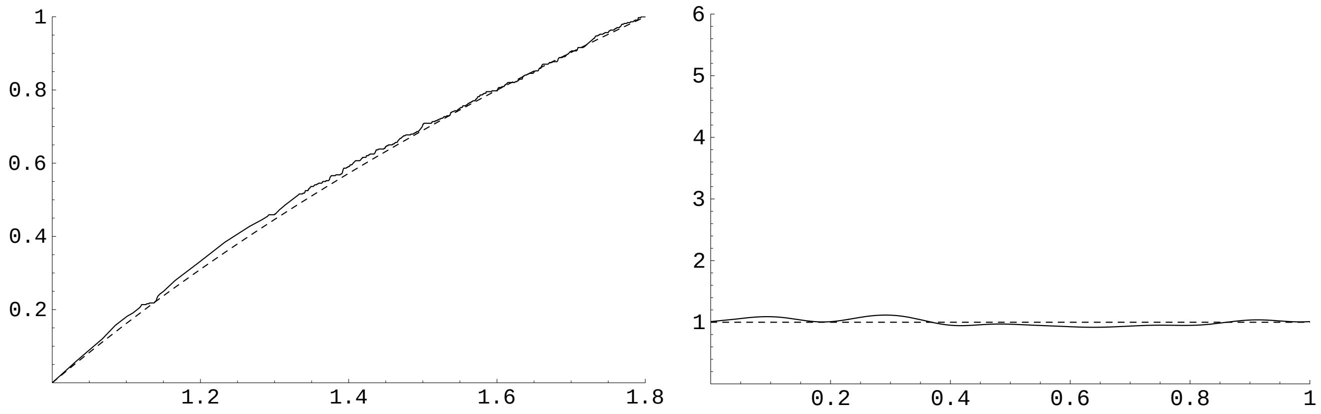


Figure 14: The Dutch municipalities: On the left the cumulative distribution functions of the mantissae (solid line) and $F_{BL_{1.8}}$ (dotted line) and on the right the density functions of the logarithms of the mantissae relative to the real base 1.8.

9.1 SAMPLING FUNCTIONS

Let $P_X \in \mathcal{P}^+_{a.c.}$. We have seen in Lemma 6.8 that $f_{\log_b(X) \pmod 1}$ is for almost all $x \in [0,1)$ a sum of functions:

$$f_{\log_b(X) \pmod 1}(x) = \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k).$$

Let us define the function $\text{sampling}_1(x,z) : \mathbb{R} \to \mathbb{R}$, with $x \in [0,1)$ fixed, as

$$\text{sampling}_1(x,z) = \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k) \cdot 1_{\{z \in [x+k, x+(k+1))\}}.$$

The subscript 1 in $\text{sampling}_1(x,z)$ indicates that the steps, the intervals where the function is constant, have length 1. For $f_{\log_b(X) \pmod 1}(x)$ one can write:

$$f_{\log_b(X) \pmod 1}(x) = \int_{\mathbb{R}} \text{sampling}_1(x,z) d\lambda(z).$$

We see that $f_{\log_b(X) \pmod 1}(x)$ is an approximation of the Lebesgue integral of $f_{\log_b(X)}$, which is equal to 1, by the Lebesgue integral of sampling_1. From this one can conclude that a distribution satisfies Benford's Law in base b if and only if for almost all $x \in [0,1)$ these approximations are perfect or

$$\int_{\mathbb{R}} \text{sampling}_1(x,z) d\lambda(z) = 1 \quad \text{for a.e. } x \in [0,1).$$

The next theorem describes what happens if the base in which one calculates the mantissae changes.

Theorem 9.1 (A change of base) Let $P_X \in \mathcal{P}^+_{a.c.}$. Let c be a strictly positive real number. We have

$$f_{\log_{b^c}(X) \pmod 1}(x) = \sum_{k=-\infty}^{\infty} f_{\log_{b^c}(X)}(x+k) = \sum_{k=-\infty}^{\infty} c \cdot f_{\log_b(X)}(c \cdot (x+k)) \qquad \text{for a.e. } x \in [0,1)$$

Proof The first equality follows from Lemma 6.8. Furthermore, we have according to Lemma 6.6 that

$$f_{\log_{b^c}(X)}(x) = \log_e(b^c) \cdot (b^c)^x \cdot f_X((b^c)^x) =$$

$$c \cdot \log_e(b) \cdot b^{c \cdot x} \cdot f_X(b^{c \cdot x}) = c \cdot f_{\log_b(X)}(c \cdot x) \qquad \text{for a.e. } x \in \mathbb{R}.$$

From this follows the second equality.

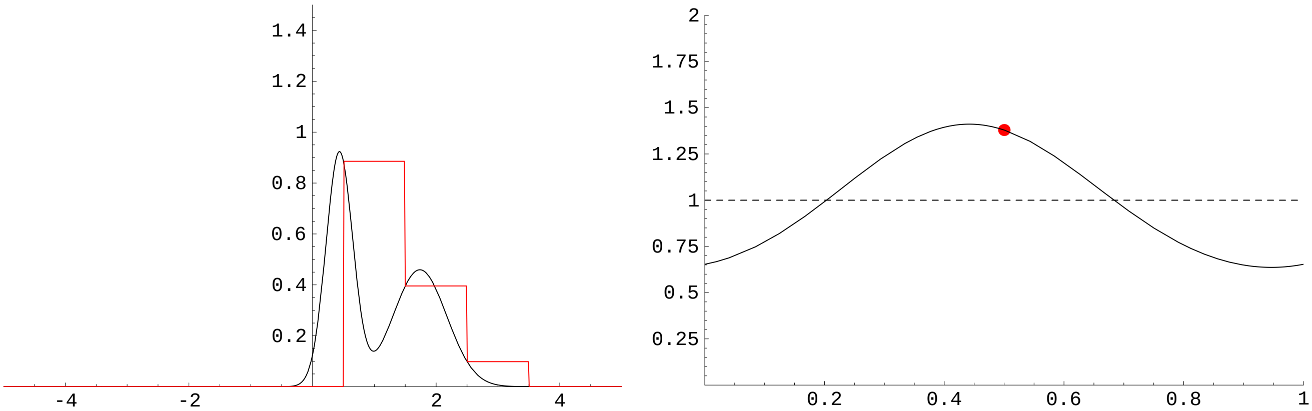


Figure 15: On the left $f_{\log_{10}(X)}(z)$ and $\text{sampling}_1(0.5, z)$ and on the right $f_{\log_{10}(X) \pmod 1}(x)$, with $f_{\log_{10}(X) \pmod 1}(0.5)$ indicated by a dot.

□

With the help of the last theorem we see that $f_{\log_{b^c}(X) \pmod 1}(x)$ is an approximation of the Lebesgue integral $\int_{\mathbb{R}} f_{\log_b(X)}(z)d\lambda(z)$, which is equal to 1, by the Lebesgue integral of the sampling function $\text{sampling}_c(x,z) : \mathbb{R} \to \mathbb{R}$ defined by

$$\text{sampling}_c(x,z) = \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(c \cdot (x+k)) \cdot 1_{\{z \in [c\cdot(x+k), c\cdot(x+(k+1))\}}$$

where the subscript c in $\text{sampling}_c(x,z)$ indicates that the steps have length c. For a.e. $x \in [0,1)$ we can write:

$$f_{\log_{b^c}(X) \pmod 1}(x) = \int_{\mathbb{R}} \text{sampling}_c(x,z)d\lambda(z).$$

We see that a distribution satisfies Benford's Law in base b^c if and only if for a.e. $x \in [0,1)$ the Lebesgue integral of the sampling function $\text{sampling}_c(x,z)$ gives a perfect approximation of the Lebesgue integral of $f_{\log_b(X)}(z)dz$, or

$$\int_{\mathbb{R}} \text{sampling}_c(x,z)d\lambda(z) = 1 \quad \text{for a.e. } x \in [0,1).$$

From this we conclude the following: The degree to which the sampling functions with step lengths c approximate the area under the graph of $f_{\log_e(X)}$ determines the degree to which distribution P_X satisfies Benford's Law in base e^c. As one expects that the area under the graph of the function $f_{\log_e(X)}$ will be approximated better by sampling functions as the steps get smaller, one expects that the distribution satisfies Benford's Law better as the base gets smaller.

Example 9.2 Let us look at a distribution with density function given by

$$f_X(x) = \frac{1}{2} \cdot \frac{1}{\sigma_1 \cdot \sqrt{2\pi} \cdot x} \cdot e^{-(\log_e(x)-\mu_1)^2/(2\cdot\sigma_1{}^2)} + \frac{1}{2} \cdot \frac{1}{\sigma_2 \cdot \sqrt{2\pi} \cdot x} \cdot e^{-(\log_e(x)-\mu_2)^2/(2\cdot\sigma_2{}^2)}$$

where $\mu_1 = 1$, $\sigma_1 = 0.5$, $\mu_2 = 4$ and $\sigma_2 = 1$. The distribution is the average of two logarithmic normal distributions. So, we observe that $P_X \in \mathcal{P}^+_{a.c.}$. In the next two images, Figure 15, we see on the left $f_{\log_{10}(X)}(z)$ and $\text{sampling}_1(0.5,z)$ and on the right $f_{\log_{10}(X) \pmod 1}(x)$.
We calculate

$$f_{\log_{10}(X) \pmod 1}(0.5) = \int_{\mathbb{R}} \text{sampling}_1(0.5,z)d\lambda(z) \approx 1.38,$$

the area under the sampling function is bigger than 1, namely 1.38 . In the image on the right, of Figure 15 we see that the area under the sampling function $\text{sampling}_1(0.8,z)$ is smaller than 1. sampling functions with steps of length 1 approximate the area under the graph of $f_{\log_{10}(X)}(z)$ very badly and therefore

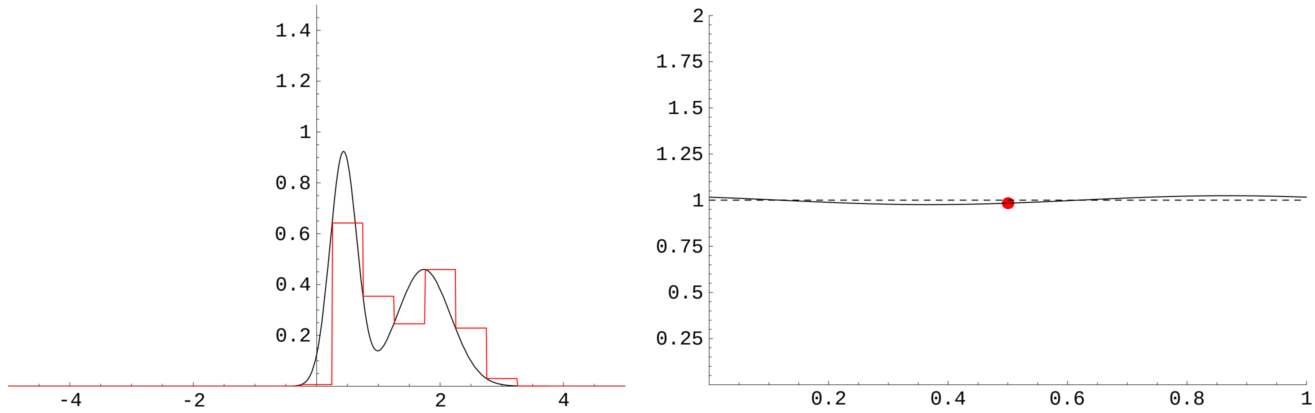


Figure 16: On the left $f_{\log_{10}(X)}(z)$ and $\text{sampling}_{\frac{1}{2}}(0.5,z)$ and on the right $f_{\log_{\sqrt{10}}(X) \pmod 1}(x)$, with $f_{\log_{\sqrt{10}}(X) \pmod 1}(0.5)$ indicated by a dot.

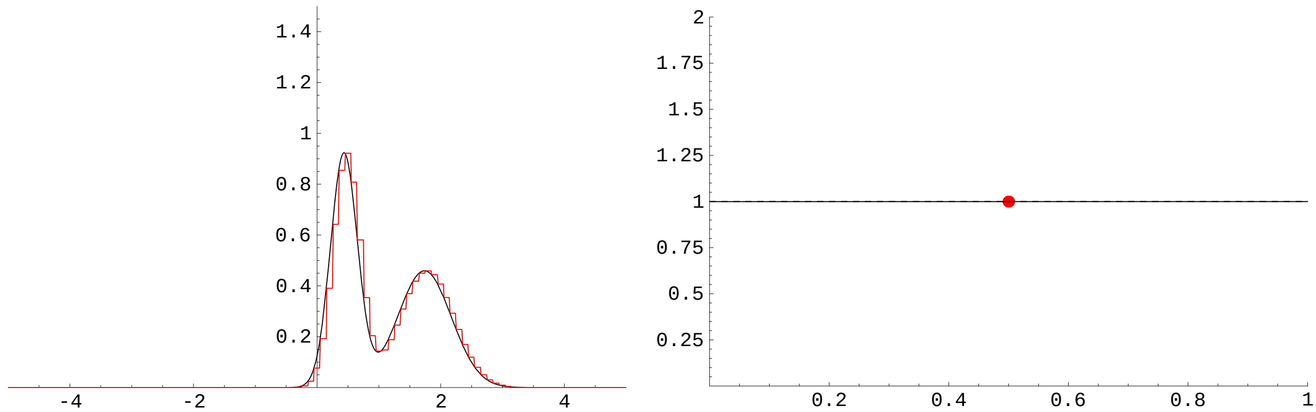


Figure 17: On the left $f_{\log_{10}(X)}(z)$ and $\text{sampling}_{\frac{1}{10}}(0.5,z)$ and on the right $f_{\log_{\sqrt[10]{10}}(X) \pmod 1}(x)$, with $f_{\log_{\sqrt[10]{10}}(X) \pmod 1}(0.5)$ indicated by a dot.

$f_{\log_{10}(X) \pmod 1}(x)$ approximates 1 very badly. Therefore the distribution P_X does totally not satisfy Benford's Law in base 10.

Let us now look what happens if we decrease the base to $b = \sqrt{10} \approx 3.16$, and plot graphs in Figure 16 analogous to the graphs plotted in Figure 15
Now we have

$$f_{\log_{\sqrt{10}}(X) \pmod 1}(0.5) = \int_{\mathbb{R}} \text{sampling}_{\frac{1}{2}}(0.5,z)d\lambda(z) \approx 0.984.$$

The areas under the graphs of the sampling functions with steps of length $\frac{1}{2}$ approximate the area under the graph of $f_{\log_{10}(X)}(z)$ very well. If one calculates in base $\sqrt{10}$, then one can expect that the mantissae of numbers that follow a distribution with density function f_X will nearly follow a $\sqrt{10}$-logaritmic distribution.

Let us finally lower the base to $b = \sqrt[10]{10} \approx 1.26$.
The length of the steps that approximate the area under the graph is 10 times smaller then the length of the steps belonging to base 10. In Figure 17 we see a nearly perfect approximation.

If the base decreases continuously, we see in the three images above on the right that the dot $f_{\log_b(X) \pmod 1}(0.5)$ converges to 1 while fluctuating continuously around it. This observation induces us to examine the continuity of $f_{\log_b(X) \pmod 1}(x)$ in b and the convergence to 1 if b decreases to 1.

9.2 Continuity of the distribution of the logarithms of the mantissae

We know that a density function of $P_{\log_b(X) \pmod 1}$ can be written as an infinite sum (Lemma 6.8). So, we will need a lemma about the continuity of a limit of functions.

Lemma 9.3 Suppose $f_n \to f$ uniformly on a set E in a metric space. Let x be a limit point of E, and suppose that

$$\lim_{t\to x} f_n(t) = A_n \quad (n = 1,2,3,\ldots).$$

Then $\{A_n\}$ converges, and

$$\lim_{t\to x} f(t) = \lim_{n\to\infty} A_n.$$

If $f_n(t)$ is continuous in $t = x$ for all n, then f is continuous in x:

$$\lim_{t\to x} f(t) = \lim_{t\to x}\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty}\lim_{t\to x} f_n(t) = \lim_{n\to\infty} f_n(x) = f(x)$$

Proof See the book Principles of Mathematical Analysis [RUD], page 149.

$\square$

Now we are ready to give the theorem about continuity.

Theorem 9.4 (Continuity) Let $P_X \in \mathcal{P}^+_{a.c.}$ and such that f_X is a continuous function on $(0,\infty)$. Let $b > 1$ and suppose that

$$\sum_{k=-n}^{n} f_{\log_b(X)}(x+k)$$

converges uniformly in x on $\mathbb{R}$. Then there is a density function $f_{\log_b(X)(\text{mod } 1)}(x)$ which is continuous in x on $(0,1)$ and right-continuous in $x = 0$ and left-continuous in $x = 1$. Suppose additionally that the sum converges uniformly in b on $[r,s]$. Then, the density function is also continuous in b on (r,s) and right-continuous in r and left-continuous in s. Furthermore,

$$\lim_{b'\to b} \gamma_{b'}(P_X) = \gamma_b(P_X)$$

Proof The function f_X is continuous on $(0,\infty)$, so also $f_{\log_b(X)}$ is continuous on $\mathbb{R}$. Since $\forall k \in \mathbb{Z}$ the function $f_{\log_b(X)}(x+k)$ is continuous in x, also the finite sum $\sum_{k=-n}^{n} f_{\log_b(X)}(x+k)$ is continuous in x for all n. Since we claimed uniform convergence of $\sum_{k=-n}^{n} f_{\log_b(X)}(x+k)$, we can use Lemma 9.3 and see that $\sum_{k=-\infty}^{\infty} f_{\log_b(X)}(x+k)$ is continuous in x on $\mathbb{R}$. According to Lemma 6.8, there is a density function $f_{\log_b(X)(\text{mod } 1)}(x)$ of $P_{\log_b(X) \pmod 1}$ that is on $[0,1)$ a.e. equal to this infinite sum. From the continuity follows that there is a density function of $P_{\log_b(X) \pmod 1}$ that is on $[0,1]$ equal to this infinite sum, and thus continuous on $(0,1)$ and right-continuous in 0 and left-continuous in 1.

Let us now prove

$$\lim_{b'\to b} f_{\log_{b'}(X)(\text{mod}1)}(x) = f_{\log_b(X)(\text{mod } 1)}(x).$$

We will prove

$$\lim_{c\to 1} f_{\log_{b^c}(X)(\text{mod } 1)}(x) = f_{\log_b(X)(\text{mod } 1)}(x).$$

We have that $\sum_{k=-n}^{n} f_{\log_{b^c}(X)}(x+k)$ converges uniformly to $f_{\log_{b^c}(X)(\text{mod } 1)}(x)$ for $b^c \in [r,s]$. So, we only have to show that the terms are continuous:

$$\lim_{c\to 1} f_{\log_{b^c}(X)}(x+k) = f_{\log_b(X)}(x+k) \qquad \text{with} \quad k \in \mathbb{Z}$$

In the proof of Theorem 9.1 we saw that

$$f_{\log_{b^c}(X)}(x) = c \cdot f_{\log_b(X)}(c\cdot x)$$

so

$$\lim_{c\to 1} f_{\log_{b^c}(X)}(x+k) = \lim_{c\to 1} c\cdot f_{\log_b(X)}(c\cdot(x+k)) = \lim_{c\to 1} f_{\log_b(X)}(c\cdot(x+k)) = f_{\log_b(X)}((x+k)).$$

The last equality holds because $f_{\log_b(X)}(x)$ is continuous in $x\in\mathbb{R}$. This proves $\lim_{c\to 1} f_{\log_{b^c}(X)(\text{mod }1)}(x) = f_{\log_b(X)(\text{mod }1)}(x)$ and therefore $\lim_{b'\to b} f_{\log_{b'}(X)(\text{mod}1)}(x) = f_{\log_b(X)(\text{mod }1)}(x)$.

We now prove that

$$\lim_{b'\to b} \gamma_{b'}(P_X) = \gamma_b(P_X).$$

Let $k_b(x) = |1 - f_{\log_b(X)(\text{mod }1)}(x)|$, then $k_b(x)$ is continuous in b on $[r,s]$ and continuous in x on $[0,1]$. This is true because $f_{\log_b(X)(\text{mod }1)}(x)$ is continuous in b and x and therefore also $1 - f_{\log_b(X)(\text{mod }1)}(x)$ and $|1 - f_{\log_b(X)(\text{mod }1)}(x)|$ are continuous in b and x. Because $[0,1]\times[r,s]$ is a compact set, $k_b(x)$ is uniform continuous in both variables. Let us take a sequence $\{b_1, b_2, \ldots\}$ that converges to b, where $b_i \in [r,s]$ for all $i\in\mathbb{N}$. Let $\varepsilon > 0$. From the uniform continuity of $k_b(x)$ in $x\in[0,1]$ for all $b\in[r,s]$ follows that there is a $\delta > 0$ such that $|b'-b| < \delta \Rightarrow |k_{b'}(x) - k_b(x)| < \varepsilon$. There is a $N\in\mathbb{N}$ such that $|b_N - b| < \delta$ and thereby $|k_{b_N}(x) - k_b(x)| < \varepsilon$. We conclude that $k_{b_n}(x) \to k_b(x)$ uniformly on $[0,1]$. There is a $N'\in\mathbb{N}$ such that for all $n \geq N'$ the functions k_{b_n} are dominated by $k_b + 1$, which is Lebesgue integrable on the interval $[0,1]$. By Lebesgue's Dominated Convergence Theorem

$$\lim_{n\to\infty} \int_0^1 k_{b_n}(x)dx = \int_0^1 k_b(x)d\lambda(x).$$

Because $\{b_1, b_2, \ldots\}$ was an arbitrary sequence that converges to b, we can conclude

$$\lim_{b'\to b} \int_0^1 k_{b'}(x)dx = \int_0^1 k_b(x)d\lambda(x).$$

□

In Example 9.7 we will see that the γ_b-distance of log normal distributions is continuous in b. See also Figure 22 and Figure 23 for continuous graphs of γ_b.

9.3 Benford towards base 1

From our experiment with population numbers one could get the idea that the γ_b-distance can get as small as one wants by lowering the base b. Also in Example 9.2 we have seen that by lowering the base the approximation by sampling functions got better. Let us make a more formal statement about this:

Theorem 9.5 (Benford towards base 1) Let $P_X \in \mathcal{P}^+_{a.c.}$ and such that $f_{\log_b(X)}$ is continuous and assume that there is a Lebesgue integrable function g and a $C > 0$ such that $\text{sampling}_c(x,z) \leq g(z)$ for $0 < c < C$. Then

$$\lim_{b\downarrow 1} \gamma_b(P_X) = 0$$

Proof By definition

$$\text{sampling}_c(x,z) = \sum_{k=-\infty}^{\infty} f_{\log_b(X)}(c\cdot(x+k))\cdot 1_{\{z\in[c\cdot(x+k),c\cdot(x+(k+1)))\}}.$$

If x and z are fixed, then k depends only on c:

$$c\cdot(x+k) \leq z < c\cdot(x+k+1)$$

This implies that $z/c - x - 1 < k \leq z/c - x$. Since k is an integer, $k = \lfloor z/c - x \rfloor$. This implies that

$$\text{sampling}_c(x,z) = f_{\log_b(X)}(c \cdot (x + \lfloor z/c - x \rfloor)).$$

Notice that

$$\lim_{c \to 0} c \cdot (x + \lfloor z/c - x \rfloor) = z.$$

Since $f_{\log_b(X)}$ is continuous

$$\lim_{c \to 0} \text{sampling}_c(x,z) = \lim_{c \to 0} f_{\log_b(X)}(c \cdot (x + \lfloor z/c - x \rfloor)) = f_{\log_b(X)}(z)$$

Now we have

$$\lim_{b \downarrow 1} f_{\log_b(X) \pmod 1}(x) = \lim_{c \to 0} f_{\log_{b^c}(X) \pmod 1}(x) = \lim_{c \to 0} \int_{\mathbb{R}} \text{sampling}_c(x,z) d\lambda(z) \text{ for a.e. } x \in [0,1)$$

By Lebesgue's Dominated Convergence Theorem this is equal to

$$\int_{\mathbb{R}} \lim_{c \to 0} \text{sampling}_c(x,z) d\lambda(z) = \int_{\mathbb{R}} f_{\log_b(X)}(z) d\lambda(z) = 1$$

Now again by Lebesgue's Dominated Convergence Theorem applied to the probability measure λ on $[0,1)$, we obtain

$$\lim_{b \downarrow 1} \gamma_b(P_X) = \lim_{b \downarrow 1} \int_{[0,1)} |1 - f_{\log_b(X) \pmod 1}(x)| d\lambda(x) = \int_{[0,1)} \lim_{b \downarrow 1} |1 - f_{\log_b(X) \pmod 1}(x)| d\lambda(x) = 0$$

□

We would like to give an example in which we apply Theorem 9.4 and Theorem 9.5. In Theorem 9.4 there is a condition about uniform convergence. We first need a lemma about uniform convergence.

Lemma 9.6 Let $P_X \in \mathcal{P}^+_{a.c.}$. Assume that there is a $M \in \mathbb{N}$ such that $f_{\log_b(X)}(x)$ is non-decreasing for $x \leq -M$ and non-increasing for $x \geq M$. Then the sequence of functions

$$\sum_{k=-n}^{n} f_{\log_b(X)}(x+k)$$

converges uniformly on $[0,1]$.
If additionally the function $f_{\log_b(X)}(x)$ is continuous, then the conditions of Theorem 9.5 are satisfied.

Proof The function $f_{\log_b(X)}(x)$ is a density function, so we have $\int_{-\infty}^{\infty} f_{\log_b(X)}(x)dx = 1$. For all $\varepsilon > 0$ there is a $M' \in \mathbb{N}$ such that

$$\int_{-\infty}^{-M'} f_{\log_b(X)}(x)dx + \int_{M'}^{\infty} f_{\log_b(X)}(x)dx < \varepsilon.$$

Next, because $f_{\log_b(X)}(x)$ is decreasing for $x \geq M$, we have for all $k \geq M$ and all $x \in [0,1)$ that $f_{\log_b(X)}(x + k) \leq f_{\log_b(X)}(0+k)$ and so

$$\sum_{k=M}^{\infty} f_{\log_b(X)}(x+k) \leq \sum_{k=M}^{\infty} f_{\log_b(X)}(0+k).$$

Also, because $f_{\log_b(X)}(x)$ is decreasing for $x \geq M$, we have that

$$k \geq M+1 \Rightarrow f_{\log_b(X)}(0+k) \leq \int_{k-1}^{k} f_{\log_b(X)}(x) d\lambda(x).$$

Therefore,

$$\sum_{k=M+1}^{\infty} f_{\log_b(X)}(0+k) \leq \int_M^{\infty} f_{\log_b(X)}(x)d\lambda(x).$$

In the same way, because $f_{\log_b(X)}(x)$ is increasing for $x \leq -M$, we have for all $k \leq -M-1$ and all $x \in [0,1)$ that $f_{\log_b(X)}(x+k) \leq f_{\log_b(X)}(1+k)$ and so

$$\sum_{k=-\infty}^{-M-1} f_{\log_b(X)}(x+k) \leq \sum_{k=-\infty}^{-M-1} f_{\log_b(X)}(1+k).$$

Also, we have that

$$k \leq -M-2 \Rightarrow f_{\log_b(X)}(1+k) \leq \int_{k+1}^{k+2} f_{\log_b(X)}(x)dx.$$

Therefore,

$$\sum_{k=-\infty}^{-M-2} f_{\log_b(X)}(1+k) \leq \int_{-\infty}^{-M} f_{\log_b(X)}(x)dx.$$

Let now $\varepsilon > 0$. There is a $M' \in \mathbb{N}$ such that $\int_{-\infty}^{-M'} f_{\log_b(X)}(x)dx + \int_{M'}^{\infty} f_{\log_b(X)}(x)dx < \varepsilon$. Let $N = \max\{M, M'\} + 2$. Let $x \in [0,1)$ and let $n \geq N$, then

$$\Big|\sum_{|k|\geq n} f_{\log_b(X)}(x+k)\Big| = \sum_{|k|\geq n} f_{\log_b(X)}(x+k) = \sum_{k=-\infty}^{-n} f_{\log_b(X)}(x+k) + \sum_{k=n}^{\infty} f_{\log_b(X)}(x+k) \leq$$

$$\sum_{k=-\infty}^{-n} f_{\log_b(X)}(1+k) + \sum_{k=n}^{\infty} f_{\log_b(X)}(0+k) \leq$$

$$\int_{-\infty}^{-N+2} f_{\log_b(X)}(x)d\lambda(x) + \int_{N-1}^{\infty} f_{\log_b(X)}(x)d\lambda(x) \leq \int_{-\infty}^{-M'} f_{\log_b(X)}(x)d\lambda(x) + \int_{M'}^{\infty} f_{\log_b(X)}(x)d\lambda(x) < \varepsilon$$

Let us prove the second statement. We have to prove that there is a Lebesgue measurable function g and a $C > 0$ such that $\text{sampling}_c(x,z) \leq g(z)$ for a.e. z for a.e. $x \in [0,1)$ and $0 < c < C$.
Define $g(z)$ as:

$$g(z) = \begin{cases} f_{\log_b(X)}(z) & \text{for } z < -M-1 \\ \max\{f_{\log_b(X)}\} & \text{for } z \in [-M-1, M+1] \\ f_{\log_b(X)}(z-1) & \text{for } z > M+1. \end{cases}$$

This function is Lebesgue integrable. Now $\text{sampling}_c(x,z) \leq g(z)$ for a.e. z for a.e. $x \in [0,1)$ and $0 < c < 1$.

□

Example 9.7 Consider a log normal distribution with a density function given by

$$f_X(x) = \frac{1}{\sigma \cdot \sqrt{2\pi} \cdot x} \cdot e^{-(\log_e(x)-\mu)^2/(2\cdot\sigma^2)} \cdot 1_{\{x>0\}}$$

Let us verify the conditions of Theorem 9.4 and Theorem 9.5. The function f is continuous on $(0,\infty)$. The function $f_{\log_b(X)}(x)$ has a bell-shaped graph, so we can use Lemma 9.6 to conclude that $\sum_{k=-n}^{n} f_{\log_b(X)}(x+k)$ converges uniformly in x on $[0,1]$, where $b > 1$. Now we have to prove that $\sum_{k=-n}^{n} f_{\log_b(X)}(x+k)$ converges uniformly in b on $[r,s]$, where $s > r > 1$. Let $b \in [r,s]$, then there is a $c \geq 1$ such that

$$\sum_{|k|>n} f_{\log_b(X)}(x+k) = \sum_{|k|>n} f_{\log_{r^c}(X)}(x+k) = \sum_{|k|>n} c \cdot f_{\log_r(X)}(c \cdot (x+k))$$

This last equality follows from Theorem 9.1. Now we will prove that for $|x|$ big enough we have that $c \in [1, \frac{\log_e(s)}{\log_e(r)}] \Rightarrow c \cdot f_{\log_r(X)}(c \cdot x) \leq f_{\log_r(X)}(x)$.

$$c \cdot f_{\log_r(X)}(c \cdot x) = c \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-(c \cdot x \cdot \log_e(r) - \mu)^2/(2 \cdot \sigma^2)} =$$

$$\frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\left(-(c \cdot x \cdot \log_e(r))^2 + 2c \cdot x \cdot \log_e(r) \cdot \mu - \mu^2 + \log_e(c) \cdot 2 \cdot \sigma^2\right)/2 \cdot \sigma^2} \leq$$

$$\frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\left(-(x \cdot \log_e(r))^2 + 2 \cdot x \cdot \log_e(r) \cdot \mu - \mu^2\right)/2 \cdot \sigma^2} = f_{\log_r(X)}(x)$$

The last inequality is true because for big $|x|$ the term $-(x \cdot \log_e(r))^2$ is the dominating term in the exponent. This $|x|$ does not depend on c. Now we can conclude that for big enough n we have for all $\frac{\log_e(s)}{\log_e(r)} \geq c \geq 1$

$$\sum_{|k|>n} c \cdot f_{\log_r(X)}(c \cdot (x+k)) \leq \sum_{|k|>n} f_{\log_r(X)}(x+k).$$

Since we know that $\sum_{k=-n}^{n} f_{\log_r(X)}(x+k)$ converges, we can conclude that $\sum_{k=-n}^{n} f_{\log_r(X)}(x+k)$ converges uniformly in b on $[r,s]$. The conditions of Theorem 9.4 are verified. The conditions of Theorem 9.5 are also satisfied: the continuity of f_X implies the continuity of $f_{\log_b(X)}$, now it follows from Theorem 9.6. We conclude that the $\gamma_b(P_X)$ varies continuously and goes to 0 if b goes to 1.

Let us examine a log normal distribution with $\mu = 0$ and $\sigma = 0.5$. For several bases we calculate the γ_b-distance and plot the graph in Figure 18. **Attention: the γ_b-distance is plotted against the natural logarithm of the base, so n means $b = e^n$.** The horizontal axis indicates the length of the steps of the sampling function that approximates $f_{\log_b(X)}$. If we want to know the γ_b-distance for base 10, then we look in the graph at $\log_e(10) \approx 2.3$, and we read that $\gamma_{10}(P_X) \approx 0.5$. We see that in base 10 the steps the length of 2.3 do not approximate very well and that the total difference between the percentages for the first significant digits and the percentages in the table given by Newcomb lies between 0 and 50. Further we derive from this graph that the discrepancy of the logarithmic distribution is very small for bases smaller than e, increases strictly if the base increases and according to Theorem 8.11 tends to 2 as the base tends to infinity. The discrepancy for base 10 is in this case big, here the variable σ is of importance: if we multiply σ with $\log_e(10)$, then the discrepancy of the logarithmic distribution is very small for bases smaller than 10 . A multiplication of the variable σ is in this case, as is a change of base, a scalar multiplication applied to $f_{\log_b(X)}$. If we for example first duplicate σ, what decreases the discrepancy, and then square the base in which you calculate, what increases the discrepancy, the degree to which the distribution satisfies Benford's Law does not change. The variable μ, which is a scalar, has no influence at the degree to which the distribution satisfies Benford's Law.

Remark further that in accordance with Theorem 9.4 and Theorem 9.5 we see that the γ_b-distance changes continuously with respect to b and goes to 1 if b goes to 1.

Example 9.8 Let us look at a logarithmic distribution P_X with density function

$$f_X(x) = \frac{1}{x \cdot \log_e(10)} \cdot 1_{\{x \in [1,10)\}}$$

By Definition 3.2 we have that $P_X = P_{BL_{10}}$. The mantissa density function in base 10 is f_{BL_b} a.e. So, the distribution satisfies Benford's Law in base 10. Here follows a graph in which the γ_b-distance is plotted against the natural logarithm of the base, so the γ_e-distance can be found at 1. On the horizontal axis we see thus the length of the steps of the sampling function that approximates $f_{\log_e(X)}$.

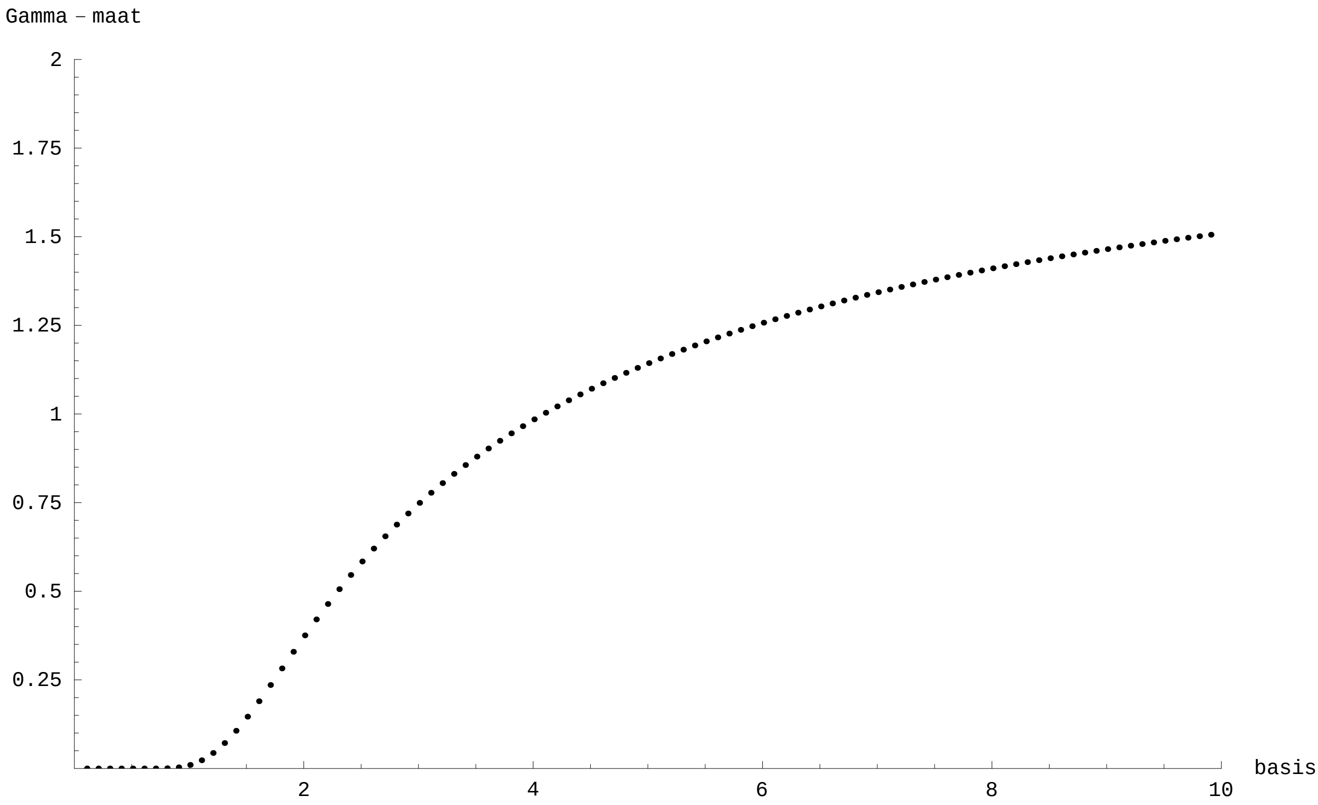


Figure 18: The γ_b-distance of a log normal distribution with $\mu = 0$ and $\sigma = 0.5$ plotted against the natural logarithm of the base b, so the γ-distance relative to base b can be found at $\log_e(b)$.

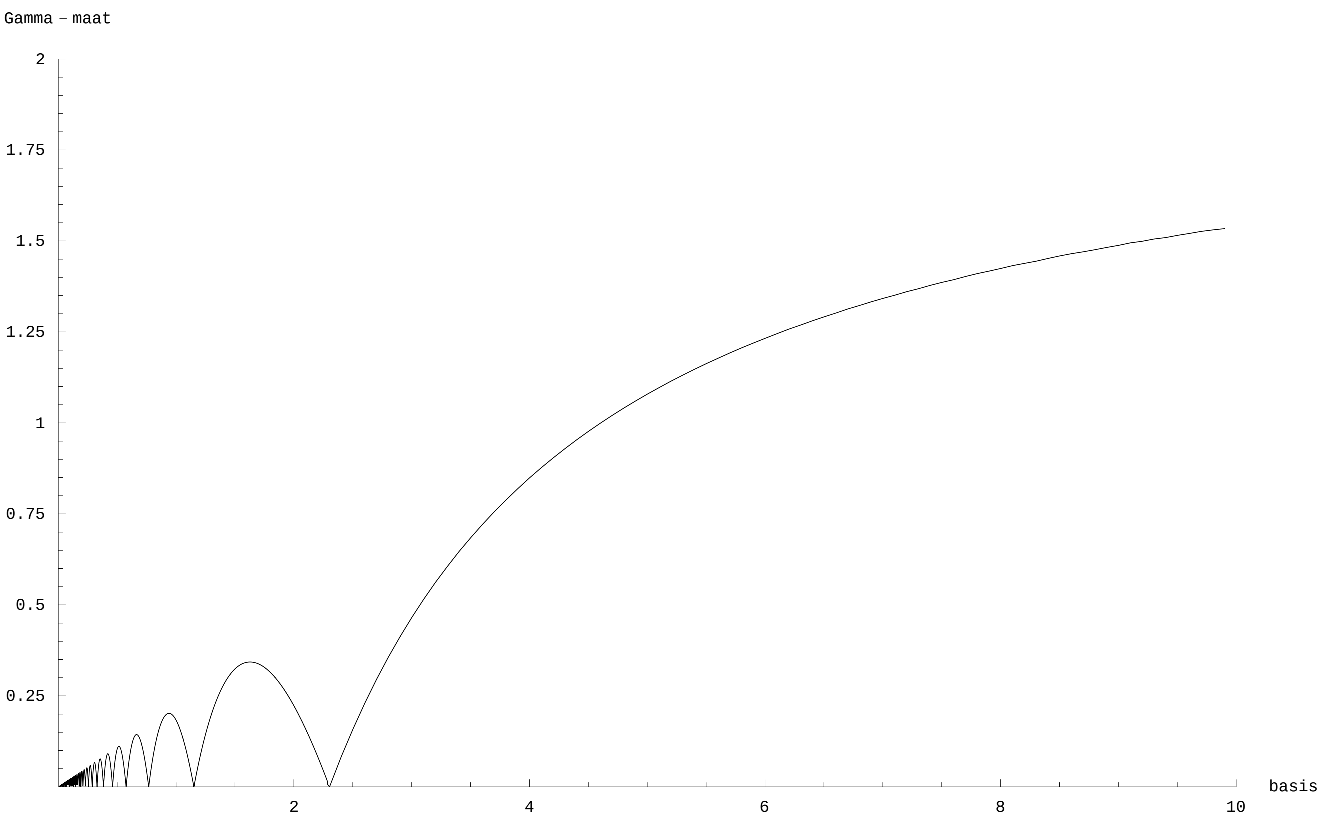


Figure 19: The γ_b-distance of Benford's Law relative to base 10, notation: $\gamma_b(P_{BL_{10}})$, plotted against the natural logarithm of the base b, so the γ_b-distance can be found at $\log_e(b)$.

We see that the discrepancy at $\log_e(10) \approx 2.3$ is zero, which means that $\gamma_b(P_X) = 0$ for $b = 10$. We conclude that Benford's Law relative to base 10 satisfies Benford's Law in base 10. Furthermore the discrepancy is zero at $\frac{1}{n} \cdot \log_e(10)$, where $n \in \mathbb{N}$, this means that the γ-distance is also zero for bases that are a n-the root of 10. However, we see that $\gamma_b(P_X)$ is non-zero between these roots, which means that Benford's Law relative to base 10 does not satisfy Benford's Law in these bases. This shows that one cannot say: this distribution satisfies Benford's Law. One can only say: this distribution satisfies Benford's Law in base b. That is why we defined satisfying Benford's Law relative to a base (Definition 5.2)

Next, it seems that in the image the γ_b-distance decreases linearly as the base tends to 1, which means that the γ_b-distance decreases logarithmically since the horizontal axes is scaled logarithmically. Maybe this linear decrease in the image is not very strange: if the length of the approximating steps gets c times smaller, then the discrepancy gets c times smaller. Finally of course the γ_b-distance goes to 2 if b goes to infinity. What about continuity? The conditions of Theorem 9.4 are not satisfied totally, the density function is not continuous in $x = 1$ and $x = 10$. Neither the conditions of Theorem 9.5 are fulfilled. Nevertheless we see that the γ_b-distance decreases continuously to 1.

From Example 9.7 and Example 9.8 it follows that by changing the base in which you calculate the γ_b-distance, the γ_b-degree to which the distribution satisfies Benford's Law can change a lot. If one wants to say something about the degree to which a distribution satisfies Benford's Law, then one has to take in account all bases $b > 1$. If one calculates the mantissa of numbers that follow the distribution of example 9.7 in base 10 then one expects a discrepancy of 50% of Simon Newcomb's table. For numbers that follow the distribution of Example 9.8 the discrepancy is 0%. It is tempting to suppose that the numbers if the second distribution do satisfy Benford's Law and the numbers of the first distribution do not. However, if one considers the first digits in another base, for example base e, then one concludes exactly the opposite. We conclude that statements about the degree to which numbers satisfy Benford's Law should be made with care. Also we showed that the γ_b-distance is not a good distance to order distributions on the degree to which they satisfy Benford's Law relative to all bases $b > 1$. The γ_b-distance only takes in account one base.

10 A FUNCTION ON DISTRIBUTIONS

If one asks oneself the question which functions are approximated well by sampling functions, one could answer that that is the case when this function does not vary too much. Let us define a function that assigns to a distribution in $\mathcal{P}^+_{a.c}$ the total variation of $f_{\log_b(X)}$. We hope that this function gives a good indication about the γ_b-degree to which a distribution satisfies Benford's Law.

Definition 10.1 Define the function $\beta_b : \mathcal{P}^+_{a.c.} \to \mathbb{R}^+$ such that it assigns to a distribution $P_X \in \mathcal{P}^+_{a.c.}$ the **total variation** of $f_{\log_b(X)}$ on $\mathbb{R}$, that is

$$\beta_b(P_X) = \lim_{N\to\infty} \sup_{W\in\mathcal{W}} \sum_{j=0}^{n_W-1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)|,$$

where the supremum is taken over the set $\mathcal{W} = \Big\{ W \text{ is a partition of } [-N,N] \text{ formed by} \{x_0, x_1, \ldots x_{n_W}\} \Big\}$, where the elements x_i of $\{x_0, x_1, \ldots x_{n_W}\}$ are such that

$$\forall \varepsilon > 0 \; \exists \delta > 0 : \; \lambda\Big(\{x \in \mathbb{R} : |x - x_i| < \delta \;\;\&\;\; |f(x) - f(x_i)| < \varepsilon\}\Big) > 0. \tag{11}$$

Defined in this way, β_b of P_X is uniquely determined. The distribution $P_{\log_b(X)}$ does not have a unique density function, that is why the elements of $\{x_0, x_1, \dots x_{n_W}\}$ have to satisfy (11).

The following Lemma is generally known:

Lemma 10.2 Let $P_X \in \mathcal{P}^+_{a.c.}$ and assume that $f_{\log_b(X)}$ is differentiable, that its derivative is integrable and that $f_{\log_b(X)}$ is of bounded variation. Then,

$$\beta_b(P_X) = \int_{-\infty}^{\infty} \left| \frac{d f_{\log_b(X)}(x)}{dx} \right| d\lambda(x).$$

Lemma 10.3 Let $P_X \in \mathcal{P}^+_{a.c.}$. We have the following properties.

1. The function β_b is scale-invariant: for all $c > 0$

$$\beta_b(P_X) = \beta_b(P_{cX})$$

2. For all $c > 0$:

$$\beta_{b^c}(P_X) = c \cdot \beta_b(P_X)$$

3. In addition let $P_{X_k} \in \mathcal{P}^+_{a.c.}$ with $1 \le k \le n$ such that they have continuous density functions $f_{\log_b(X_k)}$. Furthermore assume $f_X = \frac{1}{n} \cdot \sum_{k=1}^n f_{X_k}$. The β-value of the combination distribution P_X is smaller or equal to the mean β-value of the separate distributions:

$$\beta_b(P_X) \le \frac{1}{n} \cdot \sum_{k=1}^{n} \beta_b(P_{X_k}).$$

Proof Property 1.: According to Theorem 7.2 (2) by applying a scalar multiplication on X, a density function of $P_{\log_b(X)}$ will shift horizontally. This means that the vertical distances do not change.

Property 2.: In the proof of Theorem 9.1, we saw

$$f_{\log_{b^c}(X)}(x) = c \cdot f_{\log_b(X)}(c \cdot x) \qquad \text{for a.e. } x \in \mathbb{R}.$$

Observe that the total variation of $f_{\log_b(X)}(x)$ on $\mathbb{R}$ is the same as the total variation of $f_{\log_b(X)}(c \cdot x)$ on $\mathbb{R}$. This is true because the total variation does not change if one stretches out a function in horizontal direction. If one multiplies a function by $c > 0$, then the total variation is multiplied by c. So, the total variation of $c \cdot f_{\log_b(X)}(c \cdot x)$ is c times larger than the total variation of $f_{\log_b(X)}(x)$. This means that the total variation of $f_{\log_{b^c}(X)}(x)$ is c times larger than the total variation of $f_{\log_b(X)}(x)$.

Property 3.: Let $N \in \mathbb{R}$ and $W \in \mathcal{W}$ be arbitrary. Consider arbitrary x_j and x_{j+1} as in the definition of the β-distance. Then

$$|f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| = |\frac{1}{n} \cdot \sum_{k=1}^{n} f_{\log_b(X_k)}(x_{j+1}) - \frac{1}{n} \cdot \sum_{k=1}^{n} f_{\log_b(X_k)}(x_j)| =$$

$$\frac{1}{n} \cdot |\sum_{k=1}^{n} \left(f_{\log_b(X_k)}(x_{j+1}) - f_{\log_b(X_k)}(x_j) \right)| \le \frac{1}{n} \cdot \sum_{k=1}^{n} \left| f_{\log_b(X_k)}(x_{j+1}) - f_{\log_b(X_k)}(x_j) \right|$$

Now we also have

$$\sum_{j=0}^{n_W - 1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| \le \sum_{j=0}^{n_W - 1} \frac{1}{n} \cdot \sum_{k=1}^{n} \left| f_{\log_b(X_k)}(x_{j+1}) - f_{\log_b(X_k)}(x_j) \right| =$$

$$\frac{1}{n}\cdot\sum_{k=1}^{n}\sum_{j=0}^{n_W-1}\left|f_{\log_b(X_k)}(x_{j+1})-f_{\log_b(X_k)}(x_j)\right|.$$

We chose W arbitrarily, so we have

$$\sup_{W\in\mathcal{W}}\sum_{j=0}^{n_W-1}|f_{\log_b(X)}(x_{j+1})-f_{\log_b(X)}(x_j)|\leq\sup_{W\in\mathcal{W}}\frac{1}{n}\cdot\sum_{k=1}^{n}\sum_{j=0}^{n_W-1}\left|f_{\log_b(X_k)}(x_{j+1})-f_{\log_b(X_k)}(x_j)\right|=$$

$$\frac{1}{n}\cdot\sum_{k=1}^{n}\sup_{W\in\mathcal{W}}\sum_{j=0}^{n_W-1}\left|f_{\log_b(X_k)}(x_{j+1})-f_{\log_b(X_k)}(x_j)\right|$$

Now we let N tend to infinity to obtain:

$$\lim_{N\to\infty}\sup_{W\in\mathcal{W}}\sum_{j=0}^{n_W-1}|f_{\log_b(X)}(x_{j+1})-f_{\log_b(X)}(x_j)|\leq\lim_{N\to\infty}\frac{1}{n}\cdot\sum_{k=1}^{n}\sup_{W\in\mathcal{W}}\sum_{j=0}^{n_W-1}\left|f_{\log_b(X_k)}(x_{j+1})-f_{\log_b(X_k)}(x_j)\right|=$$

$$\frac{1}{n}\cdot\sum_{k=1}^{n}\lim_{N\to\infty}\sup_{W\in\mathcal{W}}\sum_{j=0}^{n_W-1}\left|f_{\log_b(X_k)}(x_{j+1})-f_{\log_b(X_k)}(x_j)\right|$$

□

The first property states that β_b is scale-invariant. As long as the γ-distance is scale-invariant, every function used to give an indication about the degree to which a distribution satisfies Benford's Law should satisfy this property.

According to the second property, if one uses β_b to order distributions, then also the base in which you calculate is taken into account. Squaring the base will duplicate the β_b-value, what corresponds to the duplicating of the length of the steps that approximate the area under $f_{\log_b(X)}(x)$. We already saw in figures in which the γ-distance was plotted against the base, that the base can have a big influence on the degree to which a distribution satisfies Benford's Law, see Example 10.6. However, from $\beta_e(P_X)$ one can easily calculate $\beta_b(P_X)$ for all other bases $b>1$. If one fixes a base and uses this Benford-function to order, the order will be the same for every base b. So, in the next examples we will chose $b=e$.

Example 10.4 Let us again examine log normal distributions. Let P_X be a distribution with corresponding density function given by

$$f_X(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}\cdot x}\cdot e^{-(\log_e(x)-\mu)^2/(2\cdot\sigma^2)}$$

where σ is the standard deviation and μ the mean of the normal distribution $P_{\log_e(X)}$. The total variation of $f_{\log_e(X)}$ is in this case exactly two times the maximum of $P_{\log_e(X)}$, which is $\frac{2}{\sigma\cdot\sqrt{2\pi}}$ (see Theorem 11.3). This predicts that, for a given base, this distribution will satisfy Benford's Law better as the standard deviation gets larger. If one calculates the γ-distance for several bases, then this seems to be a good prediction. We can see this in the next image: in Figure 20 the γ-distance is plotted against σ where $b=10$.

Example 10.5 In Table 4 on page 56 and page 57, we consider 9 different distributions. The function β_b has been used to order them. For each distribution we see the graphs of the distribution functions of the mantissa distributions relative to base 10, 2 and 60. The dotted lines indicate $F_{\mathrm{BL}_{10}}$, F_{BL_2} and $F_{\mathrm{BL}_{60}}$. And for each of the bases the γ-distance has been calculated.

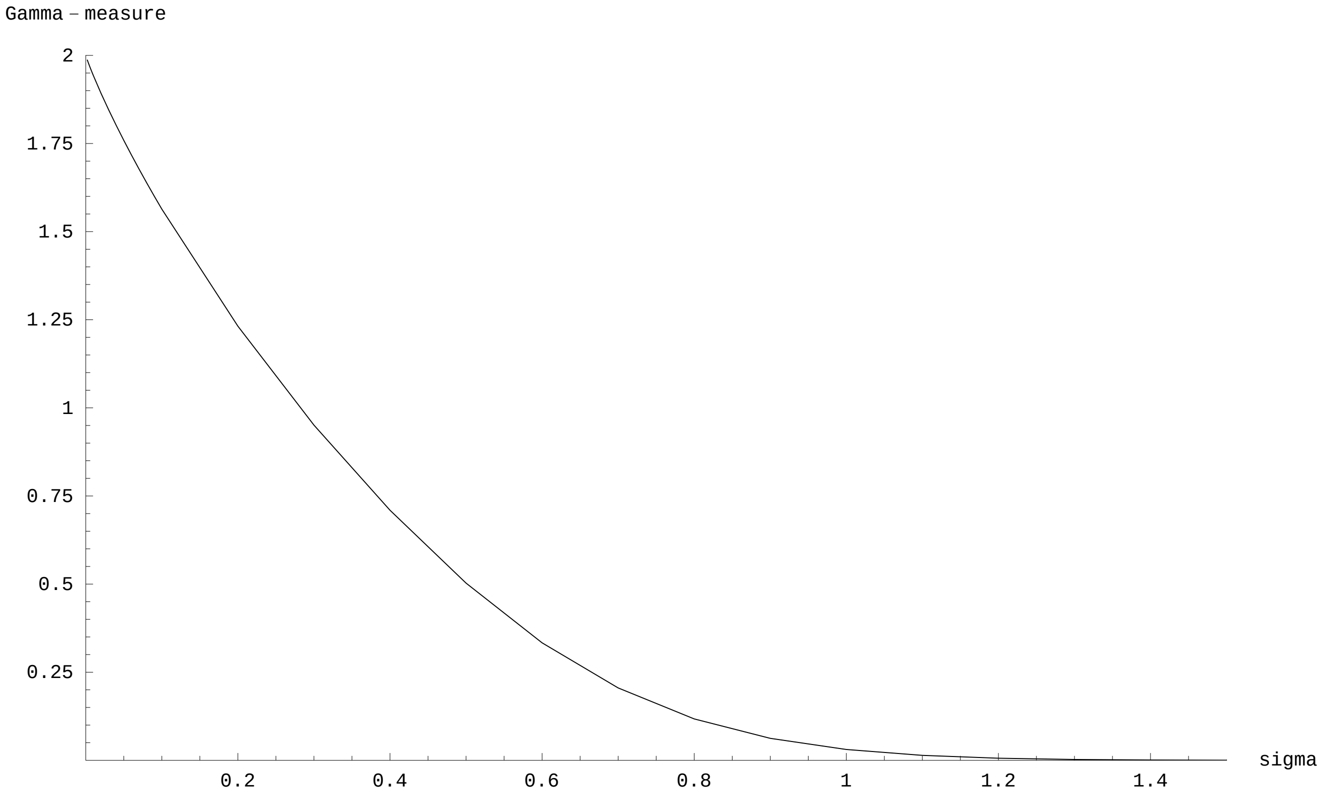
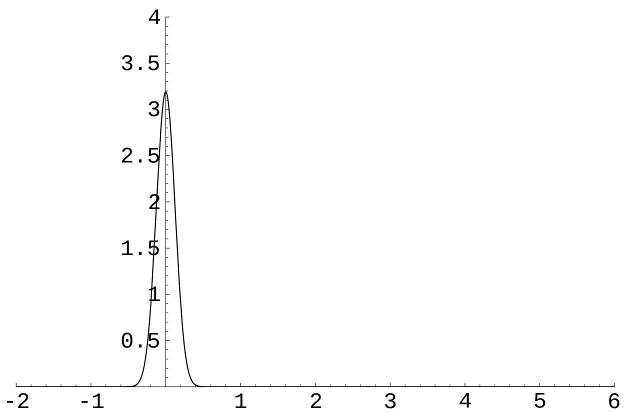


Figure 20: The γ_{10}-distance of a log normal distribution is plotted against σ.

In Table 5 on page 58 we list 19 distributions ordered by β_e- of which the γ_b-distance has been calculated in base 10, 2 and 60.

What do we see in Table 5 on page 58? First we notice that the γ_b-distance is generally increasing if we increase the base from 2 to 10 to 60. We see that β_e gives a good indication of the γ_b-distances. We see that β_e and γ_2 do not place the log-normal with $\sigma = 1$ at the same place in the ordering, but γ_{60} does places it at the same place. Concerning the three Benford-distributions: we see that the β-value of P_{60} is the lowest. So, β_e predicts that in general P_{60} is closer to Benford's Law than P_{10} and P_2. It can be seen in the table that this is a good prediction. It looks as if we found the Benford-distance we were talking about at the end of Section 3. However...



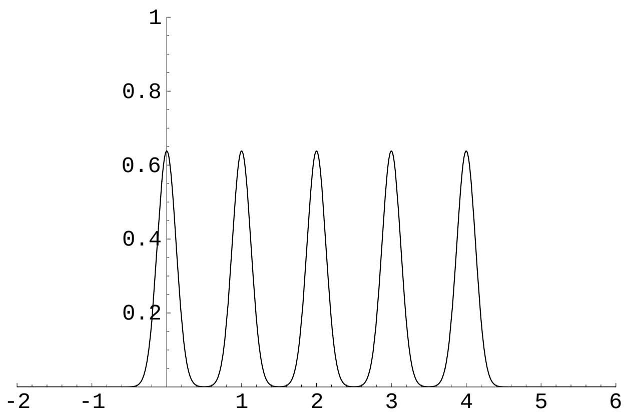


Figure 21: The graphs of the density functions of $P_{\log_e(X)}$ and $P_{\log_e(\tilde{X})}$ corresponding to Example 10.6

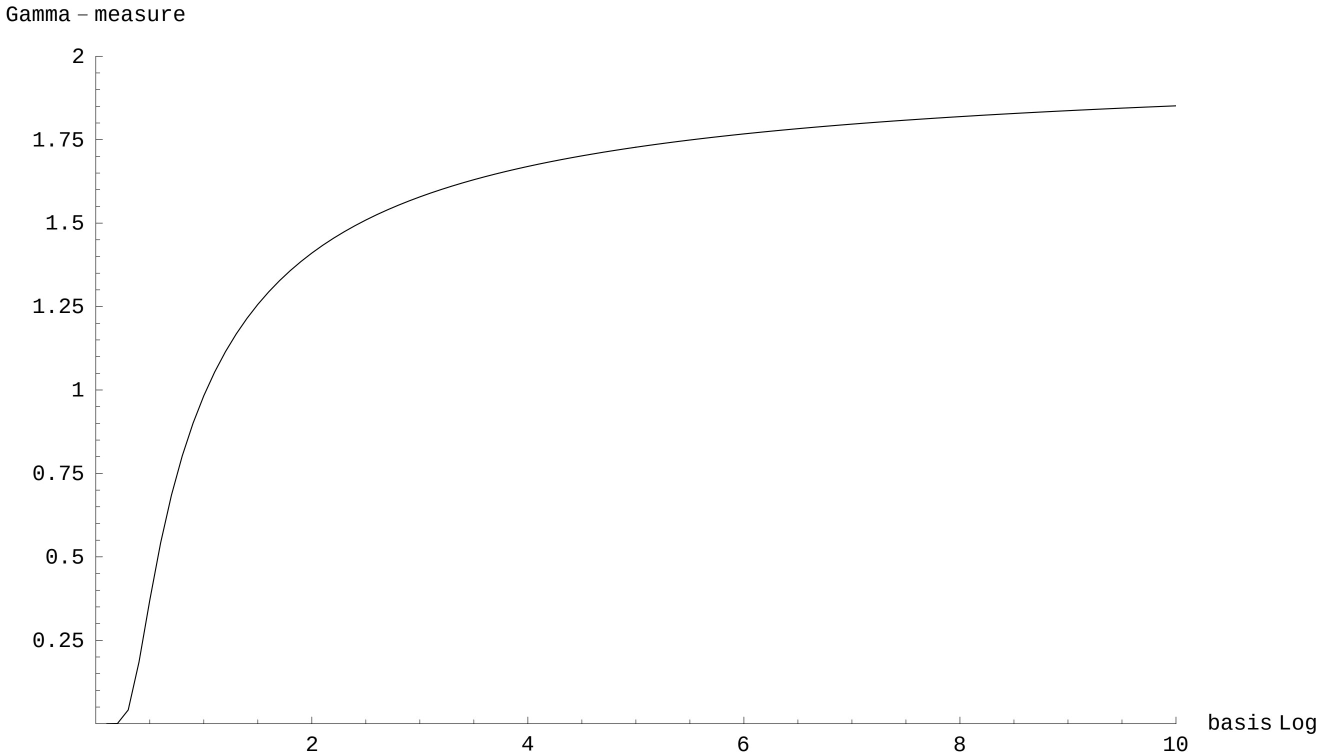


Figure 22: The γ_b-distance of a log normal distribution, $\mu = 0$, $\sigma = 1/8$, plotted against the natural logarithm of the base b on the horizontal axis, $b = e$ at 1.

Example 10.6 In this example we shall see that the total variation of $f_{\log_e(X)}$ does not always give a good indication about the γ_b-degree to which it satisfies Benford's Law. Let us compare a log normal distribution, of which the density function is given in Example 10.4 where $\mu = 0$ and $\sigma = 1/8$, with a combination of five log normal distributions, which are equal modulo scalar multiplications, a distribution $P_{\tilde{X}}$ with a density function given by

$$f_{\tilde{X}}(x) = \frac{1}{5} \cdot \sum_{i=1}^{5} \frac{1}{\sigma \cdot \sqrt{2\pi} \cdot x} \cdot e^{-(\log_e(x) - \mu_i)^2/(2 \cdot \sigma^2)} \qquad \text{for a.e. } x > 0,$$

where $\sigma = 1/8$ and $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = 2$, $\mu_4 = 3$ and $\mu_5 = 4$. Consider first in Figure 21 the graphs of the density functions $P_{\log_e(X)}$ on the left and $P_{\log_e(\tilde{X})}$ on the right. In Figure 22 and 23 the γ_b-distances of P_X and $P_{\tilde{X}}$ are plotted against the natural logarithm of the base, with e at 1. So, on the horizontal axes one reads the length of the sampling functions which approximated the area under the graph of $f_{\log_e(X)}$ and $f_{\log_e(\tilde{X})}$. Firstly remark that in both figures one sees that the γ-distance is continuous in b and tends to 0 if b tends to 1, as predicted by Theorem 9.4 and Theorem 9.5. In Figure 23 one can clearly see that the γ-distance varies a lot if the base changes. For example the distribution of the mantissae differs a lot from the logarithmic distribution in base e, where the steps are of length 1, while in base $e^{1.25}$, where the steps are of length 1.25, the mantissa distribution is nearly logarithmic. Further one can see that the γ_b-distance of P_X is for almost every base much bigger than the γ_b-distance of $P_{\tilde{X}}$, while the total variation of $f_{\log_e(X)}$ is almost the same as the total deviation of $f_{\log_e(\tilde{X})}$. For this, β is not suitable to order distributions on the degree to which they satisfy Benford's Law. It would be desirable that a Benford-distance attributes a much smaller value to this combination distribution. In the next section we will see a better way to order distributions. We will see that the total variation is useful if the density function has only one maximum. Finally remark that the γ_b-distance for $b = e$, where steps have length

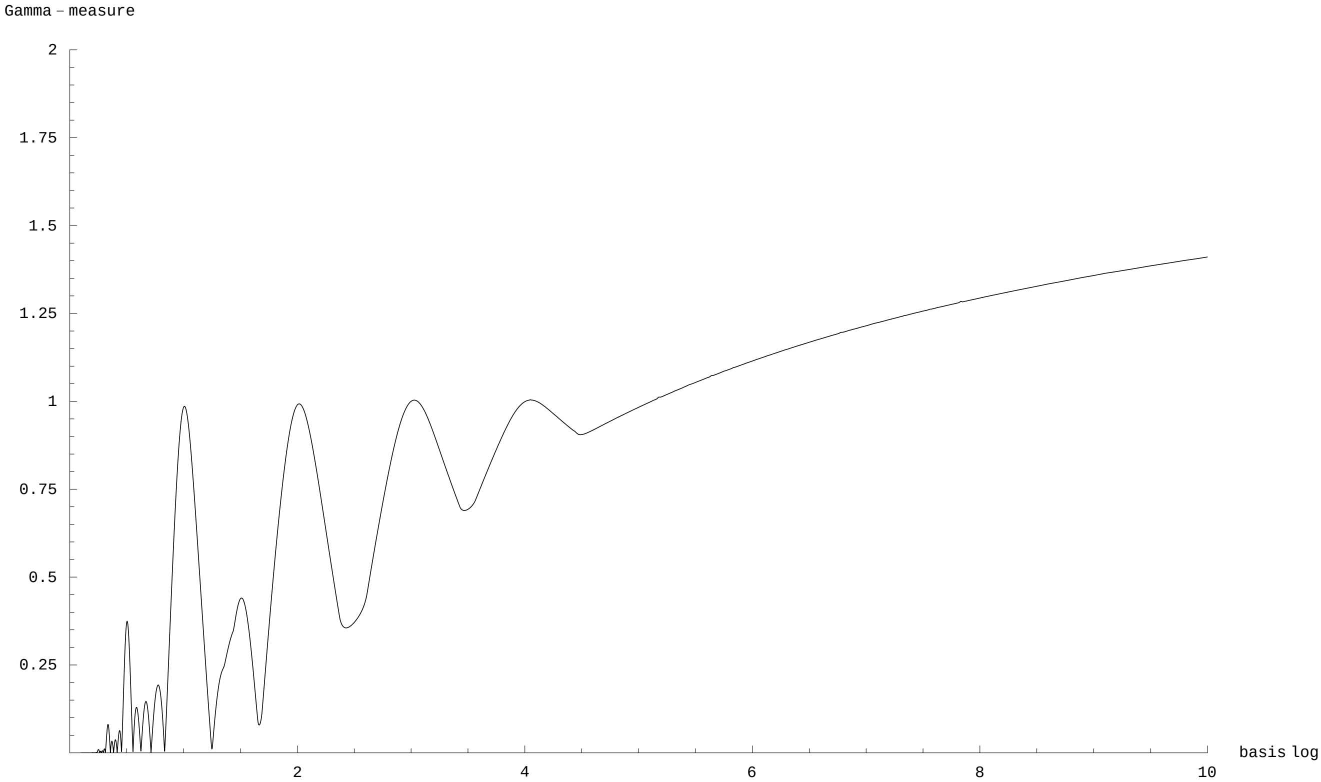


Figure 23: The γ_b-distance of a **combination** of several log normal distributions, plotted against the natural logarithm of the base b on the horizontal axis, $b = e$ at 1.

1, is the same for the log normal distribution and the combination distribution: this is the case because the distance between the peaks in the combination distribution is 1 and so the sampling functions do not notice the dales and think that they are dealing with one big peak.

11 Another function on distributions

Let us examine a function that ascribes to a distribution the essential supremum of $f_{\log_b(X)}$. Let us first give the definition of an essential supremum of a function.

Definition 11.1 Let $f : \mathbb{R} \to \mathbb{R}$. The **essential supremum** of f, with respect to Lebesgue measure, is defined as

$$\operatorname{ess\,sup} f = \inf\{a \in \mathbb{R} : \lambda(\{x \in \mathbb{R} : f(x) > a\}) = 0\}$$

Now we can give the definition.

Definition 11.2 Let $P_X \in \mathcal{P}^+_{a.c.}$. The function $\alpha_b : \mathcal{P}^+_{a.c.} \to \mathbb{R}^+$ is defined by:

$$\alpha_b(P_X) = \operatorname{ess\,sup} f_{\log_b(X)}.$$

Given $P_X \in \mathcal{P}^+_{a.c.}$, there is always a density function $f_{\log_b(X)}$ such that its supremum is equal to the α_b-value. So, from now on we will see the α_b-value of P_X as the supremum of this density function. So, we write

$$\alpha_b(P_X) = \sup_{x \in \mathbb{R}} f_{\log_b(X)}(x)$$

Theorem 11.3 Let $P_X \in \mathcal{P}^+_{a.c}$ and such that $f_{\log_b(X)}(x)$ is **unimodal**. Then,

$$\alpha_b(P_X) = \frac{1}{2} \cdot \beta_b(P_X)$$

Proof The function is unimodal, so there exist a y for which the function is non-decreasing for all $x \leq y$ and non-increasing for all $x \geq y$. The function obtains its maximum in y and therefore we have $\alpha_b(P_X) = f_{\log_b(X)}(y)$. Let $N > 0$ and let W be a partition of $[-N, y]$ formed by $\{-N = x_0, x_1, \ldots x_{n_W} = y\}$. Then,

$$\sum_{j=0}^{n_W-1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| = \sum_{j=0}^{n_W-1} f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j),$$

because $f_{\log_b(X)}(x)$ is increasing. This is a telescopic sum and thus equal to

$$f_{\log_b(X)}(y) - f_{\log_b(X)}(-N).$$

This does not depend on W, so we can conclude

$$\sup_{M \in \mathcal{M}} \sum_{j=0}^{n_W-1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| = f_{\log_b(X)}(y) - f_{\log_b(X)}(-N).$$

If we take the limit of N to infinity, we get

$$\lim_{N\to\infty} \sup_{M \in \mathcal{M}} \sum_{j=0}^{n_W-1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| = \lim_{N\to\infty} f_{\log_b(X)}(y) - \lim_{N\to\infty} f_{\log_b(X)}(-N) =$$

$$f_{\log_b(X)}(y) - 0 = f_{\log_b(X)}(y).$$

In the same way we can find

$$\lim_{N\to\infty} \sup_{M \in \mathcal{M}} \sum_{j=0}^{n_W-1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| = f_{\log_b(X)}(y),$$

where the supremum is taken over partitions of $[y, N]$. This implies that

$$\lim_{N\to\infty} \sup_{M \in \mathcal{M}} \sum_{j=0}^{n_W-1} |f_{\log_b(X)}(x_{j+1}) - f_{\log_b(X)}(x_j)| = 2 \cdot f_{\log_b(X)}(y),$$

where the supremum is taken over partitions of $[-N, N]$. So,

$$\beta(P_X) = 2 \cdot \alpha(P_X),$$

what implies the statement.

□

This theorem can be applied to all distributions listed in Example 10.5. So, α_b orders them in the same way as β_b. It also satisfies the same properties as β_b in Theorem 10.3. Let us state them and give the proofs.

Lemma 11.4 Let $P_X \in \mathcal{P}^+_{a.c.}$. We have the following properties.

1. The function α_b is scale-invariant: for all $c > 0$

$$\alpha_b(P_X) = \alpha_b(P_{cX})$$

2. For all $c > 0$:

$$\alpha_{b^c}(P_X) = c \cdot \alpha_b(P_X)$$

3. In addition let $P_{X_k} \in \mathcal{P}^+_{a.c.}$ with $1 \leq k \leq n$ and let $f_X = \frac{1}{n} \cdot \sum_{k=1}^n f_{X_k}$. The α_b-value of the combination distribution P_X is smaller or equal to the mean α_b-value of the separate distributions P_{X_k}

$$\alpha_b(P_X) \leq \frac{1}{n} \cdot \sum_{k=1}^{n} \alpha_b(P_{X_k}).$$

Proof We will first prove property (1). We have to prove that

$$\sup_{x\in\mathbb{R}} f_{\log_b(c\cdot X)}(x) = \sup_{x\in\mathbb{R}} f_{\log_b(X)}(x).$$

According to Theorem 7.2 (2) by applying a scalar multiplication on X, a density function of $P_{\log_b(X)}$ will shift. This means that the supremum does not change.
Now we will prove property (2). We have to prove that

$$\sup_{x\in\mathbb{R}} f_{\log_{b^c}(X)}(x) = c \cdot \sup_{x\in\mathbb{R}} f_{\log_b(X)}(x).$$

In the proof of Theorem 9.1 we saw that

$$f_{\log_{b^c}(X)}(x) = c \cdot f_{\log_b(X)}(c \cdot x).$$

Observe that $\sup_{x\in\mathbb{R}} f_{\log_b(X)}(x)$ is equal to $\sup_{x\in\mathbb{R}} f_{\log_b(X)}(c\cdot x)$. This is true because the supremum does not change if one stretches out a function in horizontal direction. If one multiplies a function by $c > 0$, then the supremum is multiplied by c.

We still have to prove (3). It is immediate that

$$\sup_{x\in\mathbb{R}} f_{\log_b(X)}(x) = \sup_{x\in\mathbb{R}} \frac{1}{n} \cdot \sum_{k=1}^{n} f_{\log_b(X_k)}(x) \leq \frac{1}{n} \cdot \sum_{k=1}^{n} \sup_{x\in\mathbb{R}} f_{\log_b(X_k)}(x).$$

□

Example 11.5 Let us look at a **combination** of three distributions, the combination is the average of the three distribution in the same way as in Theorem 8.20.

Distribution	$\alpha_e(P_X)$	$\gamma_{10}(P_X)$
...	...	...
Gamma(1,0.1)	*0.0755*	*0.003*
Log Normal(-6,1)	*0.3989*	*0.142*
Normal(0,1)	*0.4839*	*0.210*
...	...	...
Combination	*0.1744*	*0.064*

We read in the table that α_e-value of the combination is 0.1744, what is much lower than the mean of the three α_e-values of the separate distributions 0.3194. Also the γ_{10}-distance of the combination distribution 0.064 is smaller than the mean of the γ_{10}-distances of the three separate distributions: 0.118. Remark that in this case we looked at a combination of 3 distribution, while in newspapers the combination exists of dozens of distributions.

In the extreme case where a distribution is the combination of two distributions P_X and $P_{\tilde{X}}$ in $\mathcal{P}^+_{a.c.}$ that have mass on disjoint sets and are a scalar multiplication of each other such that $\sup_{x\in\mathbb{R}}\{f_{\log_b(X)}(x)\} =$

$\sup_{x\in\mathbb{R}}\{f_{\log_b(\tilde{X})}(x)\}$ the supremum halves. This is the case in Example 10.6 in which the supremum of the combination of the five distributions was five times smaller than the supremum of the separate distributions. So, in this example the α_b-value becomes 5 times smaller. Recall that the β_b-value did not, or nearly, change. This is the great advantage of α_b over β_b.

The supremum of $f_{\log_b(X)}$ can often give a good indication of the degree to which the distribution satisfies Benford's Law in base b. If you order distributions that are all normal distributions, all log normal distributions, all Pareto distributions, all Weibull Distributions, or all beta distributions, etc, this distance will give the right order. If for example the supremum of $f_{\log_b(X)}$ of a log normal distribution LogN(μ,σ) is lower than the supremum of $f_{\log_b(\tilde{X})}$ of another log normal distribution LogN(μ',σ'), then you will see that $\gamma_b(P_X) \leq \gamma_b(P_{\tilde{X}})$ for all real $b > 1$. Also compare Examples 6.10 and 6.11: the supremum of the density function of the standard normal distribution is more than five times bigger than the supremum of the density function of the beta distribution and indeed the latter distribution is much closer to Benford's Law. However, the supremum of $f_{\log_b(X)}$ is not the only factor that determines the degree to which the mantissae follow a logarithmic distribution: if you consider a log normal distribution of which the supremum of $f_{\log_b(X)}$ is equal to the supremum of $f_{\log_b(\tilde{X})}$ of a gamma distribution, then the mantissae of the log normal distribution will be distributed more logarithmically than the mantissae of the gamma distribution, because $f_{\log_b(X)}$ is more symmetric than $f_{\log_b(\tilde{X})}$ and therefore also easier to approximate by sampling functions. In case $f_{\log_b(X)}$ is small nearly everywhere and very big on a small set, then the α_b-value does not give a good indication of the degree to which the distribution satisfies Benford's Law. So, the function α_b, can often, but surely not always, give a good indication of the degree to which the distributions satisfy Benford's Law.

The next theorem states that the supremum of $f_{\log_e(X)}(x)$ is equal to the supremum of $x \cdot f_X(x)$.

Theorem 11.6 Let $P_X \in \mathcal{P}^+_{a.c}$. Then,

$$\alpha_e(P_X) = \sup_{x>0}\{x \cdot f_X(x)\}$$

Proof It follows easily from Lemma 6.6:

$$\alpha_e(P_X) = \sup_{x\in\mathbb{R}} f_{\log_e(X)}(x) = \sup_{x\in\mathbb{R}}\{e^x \cdot f_X(e^x)\} = \sup_{x>0}\{x \cdot f_X(x)\}$$

□

So, given a distribution in $\mathcal{P}^+_{a.c.}$, if the product of x and $f_X(x)$ is small everywhere one can expect that the γ_b-distance is small or at least smaller than the γ_b-distance of a distribution for which the supremum of the product is very big. This characterization is in agreement with the main result of the article *Survival Distributions satisfying Benford's Law* [LSE], where it is found that survival distributions satisfy Benford's Law. A survival function is a function of the form $1 - F_X$, with F_X a distribution function (for a definition see [RIC] page 348). Indeed, for this kind of functions the product $x \cdot f_X(x)$ is often small.

Finally note that a Beta distribution with a small α and a large β has a density function for which $\max_{x>0}\{x \cdot f_X(x)\}$ is small. We saw in Example 6.11 that this distribution, with small α and large β, is close to Benford's Law.

12 Conclusion

We will try to give answers to the questions formulated on page 5. We will do this in a rather intuitive way, though with the help of some theorems of the thesis.

How should we order distributions on the degree to which they satisfy Benford's Law?
It is difficult to order probability distributions on the degree to which they satisfy Benford's Law relative to all bases $b > 1$. The reason is that the γ_b-degree to which a distribution satisfies Benford's Law can change a lot while changing the base. For distributions in $\mathcal{P}_{a.c.}$ the γ_b-distance orders them correctly with respect to only one base, but the order can change if the base changes. See Example 10.6. If one wants to check if a given distribution satisfies Benford's Law, one should not check it for only one base. One can better draw a base-γ_b-distance graph. If the distribution is close to Benford's Law for all the bases around for example 10, then the statement that it satisfies Benford's Law in base 10 is more rigid.

What properties of a distribution do have great influence on the degree to which a distribution satisfy Benford's Law?
For $P_X \in \mathcal{P}^+_{a.c.}$ the α_e-distance or equivalent: $\sup_{x>0}\{x \cdot f_X(x)\}$, gives a good indication of how close to Benford's Law the distribution will be relative to all bases $b > 1$ (see Example 10.5 and Example 10.6).

Why do mantissae of numbers often follow Benford's Law?
It all is a consequence of the fact that by writing down mantissas of numbers in base b you are averaging the logarithms of the mantissae on the circle $[0,1)$. If one looks at distributions in $\mathcal{P}^+_{a.c.}$ one can see that the the function $f_{\log_b(X) \pmod 1}$ is a sum of functions on $[0,1)$ (see Lemma 6.8), summing functions can be interpreted as averaging functions.

- The lower the base, the more functions are used to sum up to $f_{\log_b(X) \pmod 1}$. In the proof of Theorem 9.1 we have seen that
$$f_{\log_{b^c}(X)}(x) = c \cdot f_{\log_b(X)}(c \cdot x).$$
So, by lowering the base, which means that c gets smaller, the function $f_{\log_{b^c}(X)}(x)$ spreads out over the real line. The more $f_{\log_{b^c}(X)}(x)$ spreads out, the more functions are used to sum up to $f_{\log_{b^c}(X) \pmod 1}$.
- Also the more distributions you mix, the more functions are used to sum up to $f_{\log_b(X) \pmod 1}$.

By averaging functions on $[0,1)$ the average of the functions gets closer to the uniform distribution, see Theorem 8.20. And if $f_{\log_b(X) \pmod 1}$ is closer to 1, then $f_{m_b(X)}$ is closer to Benford's density function f_{BL_b}, see Theorem 8.6.

13 DANKWOORD (WORDS OF THANKS IN DUTCH)

In juni 2004 kreeg ik van Karma het onderwerp voor mijn kleine scriptie: de wet van Benford. Dertig procent van de getallen zou met het cijfer 1 beginnen. In Rome, waar ik het daaropvolgende studiejaar zat, zou ik eens precies gaan uitzoeken hoe dat nou mogelijk was. Toen ik in Rome aankwam met een rugzakje met daarin wat artikelen over de wet van Benford, wist ik niet goed wat ik ermee aanmoest. Waar moest ik in hemelsnaam beginnen? Moest ik naar de kelders van het Vaticaan om daar in tabellenboekjes cijfers te gaan tellen? Dat leek mij onzin. Zo gek als Benford zelf was ik natuurlijk niet. En dus heb ik mij bezig gehouden met pizza's eten, wijn drinken en in de zon wandelen. Als ik af en toe een getal tegenkwam dat met een 1 begon, dacht ik wel: "verrek, alweer een 1!", maar echte ontdekkingen deed ik niet. Een jaar later bij terugkeer in Nederland was ik die hele wet van Benford natuurlijk straal vergeten en liep nietsvermoedend weer door het wiskundegebouw in Utrecht. Daar zag ik Karma rondlopen en dus dook ik snel een kamer in om geen verslag te hoeven doen van mijn ontdekkingen op het gebied van cijfers. Later verdween die angst en ben gewoon weer een vak gaan volgen bij Karma en in de laatste les kwam tersprake: de wet van Benford! Ik hoopte dat Karma allang vergeten was dat ik daar aan zou gaan werken, maar ze keek me aan alsof ik de grote specialist was...Thuis ben ik meteen aan het werk gegaan. Nu wist ik wel waar ik moest beginnen, ik had data nodig, wat voor een data maakte helemaal niks uit. Op internet bij een site van de overheid vond ik gegevens over het aantal inwoners van vierhonderdveertig Nederlandse gemeentes. Die begon ik ijverig één voor één over te tikken om ze in een computerprogramma te zetten. Na drie uur tikken was ik al bijna op de helft, gelukkig kwam mijn huisgenoot Bastian thuis en gaf mij de tip de kopieerfunctie te gebruiken: één minuut later waren we klaar. Hierbij zijn we aangekomen bij de eerste persoon die ik van harte wil bedanken: Bastian, bedankt! Niet alleen bedankt voor deze tip, maar ook voor het uren aan moeten horen van mijn gebrabbel over cijfers, over mijn ontdekkingen. In de weken na het binnenkomen van de eerste data viel er met mij eigenlijk niet te leven, apathisch zat ik achter mijn computer of gewoon in een stoel recht voor mij uit te staren. Na twee maanden vierentwintig uur per dag zeven dagen per week over cijfers te hebben gedacht, heeft Bastian mij meegenomen naar Italië. Na twee dagen vroeg hij of ik al een beetje ontstresst was, waarna ik een zeer vreemde schokkende beweging met mijn hoofd maakte en zei dat het wel weer ging. De pizza's wisten mij ditmaal niet te beroeren, de cijfers bleven door mijn hoofd dwalen. Weer thuis werd het tijd eens bij Karma langs te gaan. Zij ontving mij aardig in haar werkkamer en was enthousiast over mijn ideeën. Elke twee weken zou ik bij haar langskomen. Ze liet me uren op haar krijtbord krabbelen en luisterde erg goed naar alles wat ik te melden had. En dat zeker voor een paar maanden lang! Wat een geduld en wat een interesse! Mijn beste ideeën kreeg ik dan ook altijd tijdens de terugreis in de trein op de dagen dat ik bij haar was langs geweest. Zij is de tweede persoon: Karma, mag ik u zeer hartelijk bedanken voor al uw geduld en aanmoedigingen om door te gaan. Het uitschrijven van mijn ontdekkingen bleek erg moeilijk te zijn, omdat ik eigenlijk niks van kansrekening bleek te weten. Ik schreef alles in een soort zelfverzonnen wiskunde op. Na een jaar was het toch zo opgeschreven dat ik het durfde op te sturen naar experts op het gebied van de wet van Benford. Karma raadde me aan contact op te nemen met Ted Hill, een emeritus professor uit Atlanta in Georgia (VS): de grootste expert op dit gebied. Deze reageerde meteen erg aardig en was zelfs bereid om in zijn vakantie die hij in Nederland doorbracht samen met mij uren door mijn scriptie te ploegen. Dwars door mijn zelfbedachte wiskunde heen zag hij toch mijn ideeën en leerde me hoe een echte wiskundige het zou opschrijven. Beetje bij beetje werd mijn scriptie begrijpelijk. En toen Ted na een jaar weer terug in Nederland kwam, was hij tevreden over het resultaat. Hij is de derde persoon: dank u Ted voor al die uren die u mij heeft geholpen, mijn scriptie is nu leesbaar! (Voor een wiskundige althans.)

Verder dank ik de wiskundigen: Arno Berger (Canada), Laurence Leemis (VS) en Peter Schatte (Duitsland).

Ten slotte dank ik mijn vader voor alle aanmoedigingen, maar ook voor het meedenken, je had goede ideeën pappa! En niet te vergeten mijn moeder en Kirsten (voor het er altijd voor me zijn), Rutger (voor het aansporen tot studeren), Jet en Rien (voor het altijd beschikbaar stellen van hun huis om in te studeren) en alle andere mensen.

Dan nog een scriptiespreuk: “kijk om je heen en vind de 1”.

Ik groet u allen,

Jesse Dorrestijn

PS: U heeft vast opgemaakt uit het aantal bladzijden van deze sciptie dat zij van een kleine is opgewaardeerd tot een grote scriptie (Masterscriptie).

References

[BEN] Benford, F. – *the law of anomalous numbers*, Proc. Amer. Phil. Soc. **78** (1938), 551-572.

[DIA] Diaconis, P. – *The distribution of leading digits and uniform distribution mod* 1, Ann. Probab. **5** (1977), 72-81.

[HI1] Hill, T. – *Base-invariance implies Benford's law*, Proc. Amer. Math. Soc. **123** (1995), 887-895.

[HI2] Hill, T. – *A statistical derivation of the significant-digit law*, Statist. Sci. **10** (1995), 354-363.

[JR] Janvresse, E., de la Rue, T. – *From Uniform Distributions To Benford's Law*, Applied Probability Trust **41** (2004), 1203-1210.

[KNU] Knuth, D.E. – *The Art of Computer Programming*, vol 2, Addison-Wesley (1981)

[LSE] Leemis, L.M., Schmeiser B.W., Evans, D.L. – *Survival Distributions Satisfying Benford's Law*, The American Statistician **54** (2000), 236-241.

[NEW] Newcomb, S. – *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. **4** (1881), 39-40.

[RAI] Raimi, R. – *The First Digit Problem*, Am. Math. Mon. **83** (1976), 521-538.

[RIC] Rice, J.A. – *Mathematical Statistics and Data Analysis*, second edition, Duxbury Press (1995)

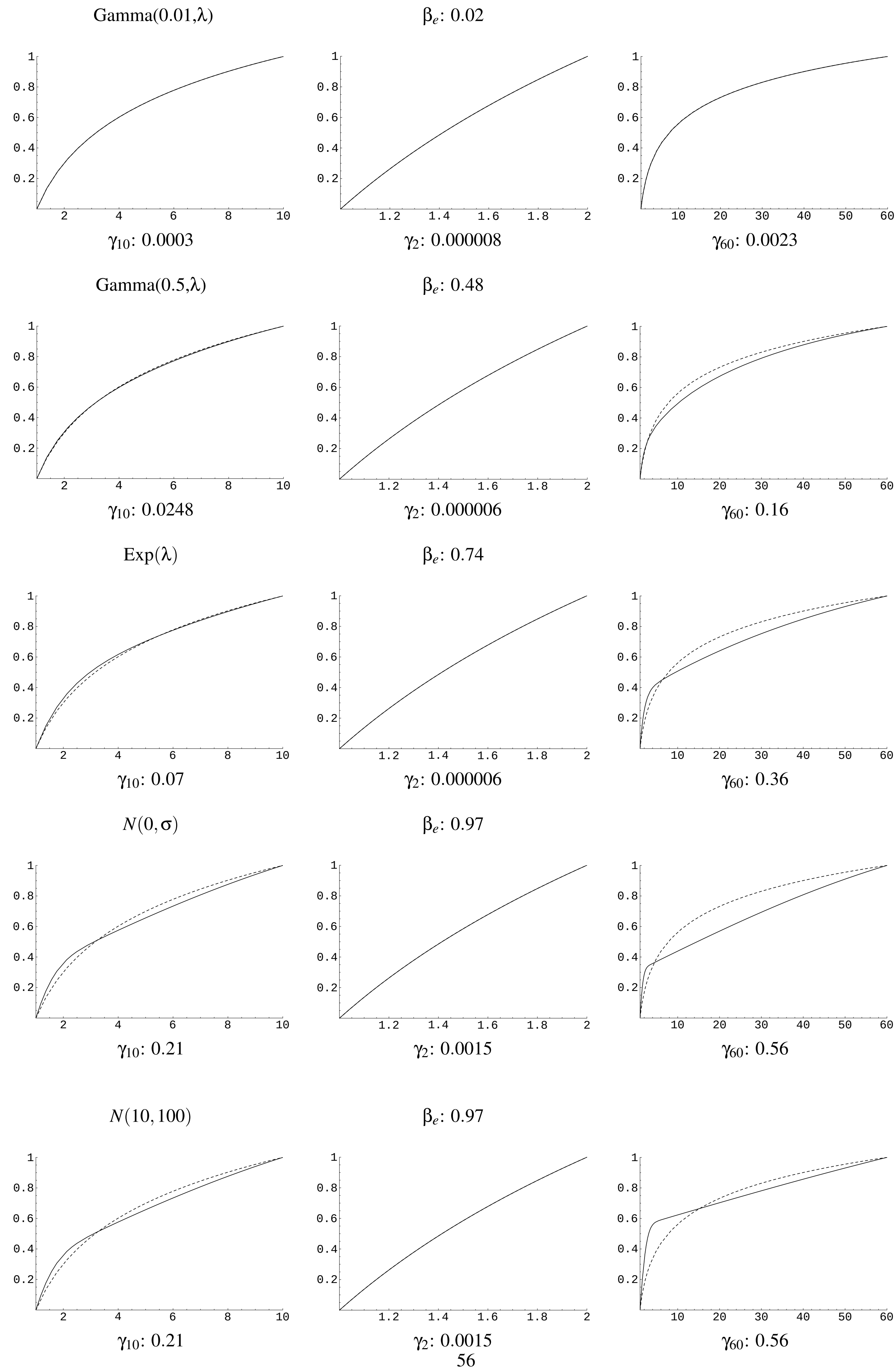
[RUD] Rudin, W. – *Principles of Mathematical Analysis*, third edition, McGraw-Hill (1953)

[SC1] Schatte, P. – *On Random Variables with Logarithmic mantissa Distrution Relative to Several Bases*, Journal of Information Processing and Cybernetics **17** (1981), 293-295.

[SC2] Schatte, P. – *On Benford's law to variable base*, Statistics and Probability Letters **37** (1998), 391-397.

[WEI] Weil, H. – *Über die Gleichverteilung von Zahlen mod. Eins*, (1916).

[WIL] Williams, D. – *Probability with Martingales* ,Cambridge University Press (1991)



Gamma(0.01,λ)
β_e: 0.02
γ_{10}: 0.0003
γ_2: 0.000008
γ_{60}: 0.0023
Gamma(0.5,λ)
β_e: 0.48
γ_{10}: 0.0248
γ_2: 0.000006
γ_{60}: 0.16
Exp(λ)
β_e: 0.74
γ_{10}: 0.07
γ_2: 0.000006
γ_{60}: 0.36
$N(0,\sigma)$
β_e: 0.97
γ_{10}: 0.21
γ_2: 0.0015
γ_{60}: 0.56
$N(10,100)$
β_e: 0.97
γ_{10}: 0.21
γ_2: 0.0015
γ_{60}: 0.56

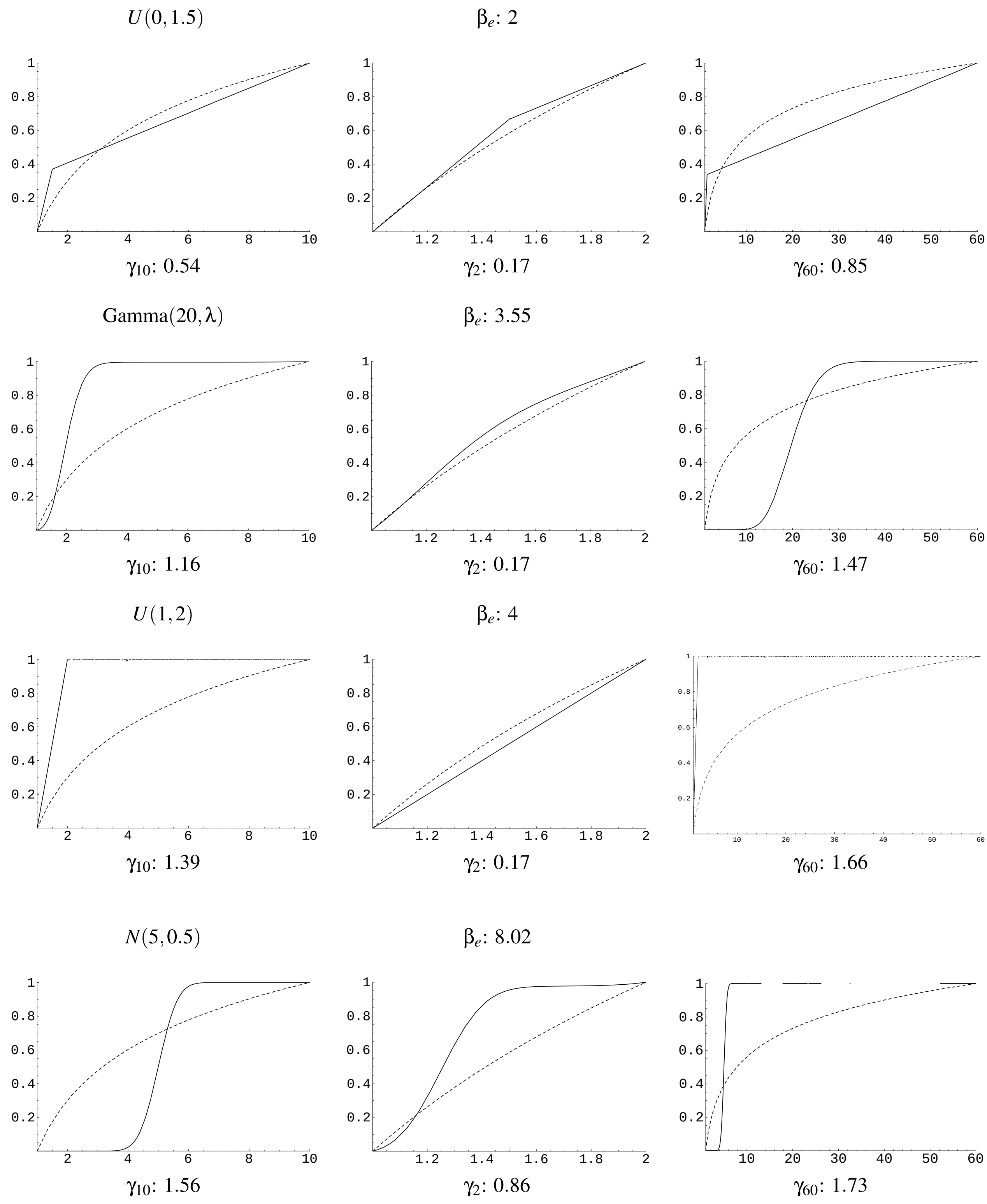


Table 4: Example 10.5: The mantissa distribution functions relative to base $b=10$, $b=2$ and $b=60$ and the γ_b-distances relative to base $b=10$, $b=2$ and $b=60$ of distributions which are ordered by β_e.

Distribution	β_e	γ_{10}	γ_2	γ_{60}
Gamma(0.01,λ)	0.0191	$2.70 \cdot 10^{-4}$	$8.80 \cdot 10^{-9}$	$2.34 \cdot 10^{-3}$
Weibull($0.5, \lambda > 0$)	0.368	$1.41 \cdot 10^{-3}$	$5.83 \cdot 10^{-12}$	$4.51 \cdot 10^{-2}$
LogNormal(μ,2^2)	0.399	$5.94 \cdot 10^{-7}$	$1.58 \cdot 10^{-16}$	$1.15 \cdot 10^{-2}$
Gamma(0.5,λ)	0.484	$2.48 \cdot 10^{-2}$	$1.18 \cdot 10^{-6}$	$1.62 \cdot 10^{-1}$
Benford$_{60}$	0.488	$1.94 \cdot 10^{-1}$	$2.86 \cdot 10^{-2}$	0
Weibull($0.8, \lambda > 0$)	0.589	$2.78 \cdot 10^{-2}$	$2.00 \cdot 10^{-7}$	$2.17 \cdot 10^{-1}$
Exp(λ)	0.736	$7.25 \cdot 10^{-2}$	$6.29 \cdot 10^{-6}$	$3.56 \cdot 10^{-1}$
LogNormal(μ,1^2)	0.798	$3.08 \cdot 10^{-2}$	$1.74 \cdot 10^{-11}$	$3.92 \cdot 10^{-1}$
Benford$_{10}$	0.869	0	$1.31 \cdot 10^{-1}$	$8.75 \cdot 10^{-1}$
$N(0, \sigma^2)$	0.968	$2.12 \cdot 10^{-1}$	$1.45 \cdot 10^{-3}$	$5.57 \cdot 10^{-1}$
$N(10, 100^2)$	0.968	$2.12 \cdot 10^{-1}$	$1.46 \cdot 10^{-3}$	$5.57 \cdot 10^{-1}$
Weibull($2.1, \lambda > 0$)	1.55	$4.79 \cdot 10^{-1}$	$7.53 \cdot 10^{-3}$	$9.31 \cdot 10^{-1}$
LogNormal(μ,0.5^2)	1.60	$5.03 \cdot 10^{-1}$	$4.41 \cdot 10^{-5}$	1.00
$U(0, 1.5)$	2	$5.38 \cdot 10^{-1}$	$1.72 \cdot 10^{-1}$	$8.47 \cdot 10^{-1}$
Benford$_2$	2.89	1.40	0	1.66
Gamma($20, \lambda$)	3.55	1.16	$1.66 \cdot 10^{-1}$	1.47
$U(1,2)$	4	1.40	$1.72 \cdot 10^{-1}$	1.66
LogNormal(μ,0.1^2)	7.98	1.56	$8.69 \cdot 10^{-1}$	1.73
$N(5, 0.5^2)$	8.02	1.56	$8.68 \cdot 10^{-1}$	1.73

Table 5: Example 10.5: γ_b-distances relative to base $b = 10$, $b = 2$ and $b = 60$ of distributions which are ordered by β_e.