Bachelor of Science in Cognitive Science

Bachelor's Thesis

On Benford's law

Computing a Bayes factor with the Savage-Dickey method to quantify conformance of numerical data to Benford's law

Nina Mogan Mainusch

First Supervisor: Prof. Dr. Michael Franke

Second Supervisor: Luke Effenberger

Universität Osnabrück

February 20, 2020

Abstract

Benford's law is a statistical phenomenon stating that the distribution of significant digits occurring in various natural data sets, amongst others population numbers, stock prices and physical constants, follows a logarithmically decaying pattern. By conducting an exploratory behavioral experiment, we confirm that this regularity is neglected to preserve when humans forge numbers, where forensic digit analysis can reveal the irregularities and support the detection of fraud in scientific, economic and political data. The classical null hypothesis significance testing approach, represented amongst others by the χ^2-test, cannot quantify how conform the data is to the Benford distribution, but returns only dichotomous decisions. With an increasing amount of data, any deviation from Benford's law is classified as significant, not allowing to specify the researcher's degree of uncertainty about the data generating process. These detriments can be addressed with the Bayesian approach to hypothesis testing, as it is proposed in this thesis. Employing the Savage-Dickey method, we calculate a Bayes factor for different prior distribution specifications in the Multinomial-Dirichlet model. By means of a simulation study, we show that the proposed prior distributions allow for a differentiated evaluation of the data and that the Bayesian approach is not prone to reject the null hypothesis that the data conforms to Benford's law disproportionately often.

Declaration of Authorship

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

City, Date

Signature

Contents

Chapter 1

Introduction

Any number is composed of digits, where a digit is an integer from the set $\{0, 1, ..., 9\}$, except for the first digit, which is by convention never 0. Intuitively, one expects that the digits of a number occur with equal frequency. The first digit should be $1, ..., 9$, with the frequency $\frac{1}{9}$ and the subsequent digits should each occur with frequencies close to $\frac{1}{10}$. Benford's law, also known as the significant-digit or first-digit law, however, is the statistical phenomenon stating that the frequencies of leading digits in natural data sets follow a logarithmic distribution. The digit 1 is the first significant digit in 30.1% of the cases, but the digit 9 in less than 5% of the cases (Tab. 1.1). These frequencies correspond to the spacing on the logarithmic number line, since the distance between 1 and 2 on the logarithmic scale is $\log(2) - \log(1) = 0.301$, the distance between 2 and 3 $\log(3) - \log(2) = 0.176$, et cetera. Formally declared, Benford's law for the first significant digit states

$$\begin{aligned} \text{Prob}(D_1 = d_1) &= \log_{10}(d_1 + 1) - \log_{10}(d_1) \\ &= \log_{10}\Big(1 + \frac{1}{d_1}\Big), \end{aligned} \tag{1.1}$$

where D_1 denotes the first significant digit of a number $x \in \mathbb{R}$ and $d_1 \in \{1, 2, ..., 9\}$. For instance $D_1(314.15) = 3$ and $D_1(0.0017724) = 1$[1].

Benford's law is applied in forensic digit analysis based on the reasoning that if a numerical data set is supposed to conform to Benford's law, a deviation

[1]By Prob we will denote the relative frequency that an event occurs. For example, $\text{Prob}(D_1 = 1) = \frac{\#1}{n}$ is the limiting proportion of the first significant digit 1, where $n \leq N$, $N \to \infty$, and #1 is the amount of numbers with leading digit 1. We assume that all the limiting proportions exist.

d_1	1	2	3	4	5	6	7	8	9
$\text{Prob}(D_1 = d)$	0.301	0.176	0.125	0.097	0.08	0.067	0.058	0.05	0.046

Table 1.1: Benford's law - Probabilities of the first digits.

from the Benford predicted frequencies indicates forgery of the data. Numerical data sets are screened for anomalies with goodness-of-fit tests, where a point null hypothesis represents conformance to Benford's law and the alternative indicates non-conformance. Null hypothesis significance testing (NHST) is the established approach to perform these tests, employing mainly the χ^2-test. Nevertheless, given measurement errors and variance in the data, no data set will conform exactly to Benford's law. The more data, the higher the probability that the NHST tests judge any deviation from Benford's law significant. As a consequence, the null hypothesis is disproportionately often rejected, leading to an increase in falsely indicated absence of Benford's law when in reality the data sets are conform to it. Furthermore, using the NHST approach does not allow to quantify evidence in favour of the null or alternative hypothesis, but instead dichotomous decisions about the conformance to Benford's law are made.

In this thesis a Bayesian approach to investigate Benford data is proposed, representing the data in a Multinomial-Dirichlet model. Instead of calculating a p-value, which is often misinterpreted and not a quantitative measure for the strength of evidence the data provides, we will compute a Bayes factor with the Savage-Dickey method. The Bayes factor is a continuous measure, allowing to gather evidence in favour of the null or in favour of the alternative hypothesis. Computed in a Bayesian hypothesis testing environment, this method is an alternative to the established goodness-of-fit tests performed in NHST.

Regarding the structure of the thesis, we will start by defining Benford's law for the first k digits, presenting its mathematical properties and expounding empirical evidence for it in chapter 2. Chapter 3 is dedicated to elaborate the difference between the Bayesian and the frequentist approach to hypothesis testing, in order to understand our choice in favour of a Bayesian model. The Bayesian model enables us to compute a Bayes factor efficiently via the Savage-Dickey method. In the fourth chapter round numbers are introduced and based on the results of an exploratory experiment that we conducted, we argue that humans have a preference for certain digits when forging numbers. The fifth chapter compares the Bayes factor to the χ^2-test results for different sample sizes and prior distribution specifications. Utilising our findings from chapter 4, we gradually distort genuine Benford data with human generated data to inspect the effects to the two hypothesis test results. Lastly, a conclusion about applying the Bayesian method instead of the conventional one is drawn and the model's limitations are discussed in chapter 6.

Chapter 2

Look under the hood

2.1 Benford's law

2.1.1 History

Following Stigler's law of eponymy – i.e. no scientific discovery is named after its original discoverer – it wasn't Benford who first discovered this phenomenon but the astronomer Simon Newcomb in 1881 [25]. He noticed "how much faster the first pages [of logarithmic tables] wear out than the last ones" and stated that "the law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally likely." The mantissa of a number $x \in \mathbb{R}$ is the decimal part of its logarithm. For example, if $x = 1234.56$ and $\log_{10}(1234.56) = 3.09151$, then the mantissa of x is 0.09151. Fifty years later, in 1938, the physicist Frank Benford rediscovered and published Newcomb's discovery as the "law of anomalous numbers", supported by over 20.000 data points from 20 different categories such as atomic weights, population data and American League baseball statistics [2].

2.1.2 Applications

It is the astonishing universal appearance of Benford's law in diverse disciplines, amongst others the field of social science, economics, natural science, medicine and computer science, that kept people fascinated throughout the last 150 years. Benford's law is used to analyse voter's behaviour and strategies in political election data [22, 30] and to detect election fraud [21, 41]. Another field of application is medicine, where it is extensively used to detect data manipulation in scientific publications [35, 39] or to predict and treat diseases [44]. Additionally, Benford's law is used for medical imaging, providing diagnostic information about potential diseases [7]. Because the Benford distribution is scale- and base-invariant, it can be used to statistically classify the image data (cf. chapter 2). It is not only medical image data that can be tested with Benford's law, but image and computer graphics data in general. When intending to prove that certain images are processed and not genuine, it can be taken advantage of the fact that light intensities in certain natural images follow Benford's law, whereas artificial images most likely do not [1]. In addition, with Benford's law one can detect and anticipate natural phenomena like earthquakes [40, 43] or the duration of volcanic eruptions [13].

Promoted by Mark Nigrini, a well-known application of Benford's law is the detection of accounting fraud [26, 27]. There are plenty of case studies about financial frauds that motivate using Benford's law for fraud detection. One of them is "Lessons from an $8 million fraud" by Mark Nigrini, which is about Nathan J. Mueller's pilfering of financial services [29]. By accident he got the authority to approve company checks of up to $250,000. He abused this authority to approve his own check requests and over the time period of four years he embezzled about $8.5 million. In the end it was not Benford's law, but an accident that got Mueller caught and he was sentenced to 97 months in prison[1]. The lesson is that there are forensic tools and tests that could have prevented or detected the fraud much earlier. One could have scanned the recently approved checks for deviations from Benford's law, for a particular increase in specific first digits or for an exponential growth in check sums. When faking checks and numbers, people start with a small amount of money and only a few checks. When they recognise that their fraud was successful, they increase the forged amounts of money exponentially, until they reach a certain threshold which cannot be exceeded, e.g. due to a limit of check approval authority.

There are real world examples where frauds where identified based on deviations from Benford's law. The Enron accounting scandal is one of those [27]. A deviation from Benford's law was also detected in the numbers of Greece's economic reports to European authorities, and Greece was indeed accused of data manipulation by the European Commission [37].

[1] More such case studies can be found in Miller's book [23].

2.1.3 Benford distributed data

It is possible to find data that follows Benford's law without much effort. Considering the first 24 numbers of the Fibonacci sequence, one can observe how often a number starts with 1 compared to 9 (Tab. 2.1). Obtaining the first digit distribution of the

1	1	2	3	5	8	13	21
34	55	89	144	233	377	610	987
1597	2584	4181	6765	10946	17711	28657	46368

Table 2.1: The first 24 Fibonacci numbers.

first 1000 numbers of the Fibonacci sequence, one can notice quite clearly how closely they conform to the frequencies predicted by Benford's law (Fig. 2.1A). Even when considering the absolute frequency of twitter followers, once can recognise Benford's law (Fig. 2.1B)[2]. Examples such as consumer product prices, numbers drawn in

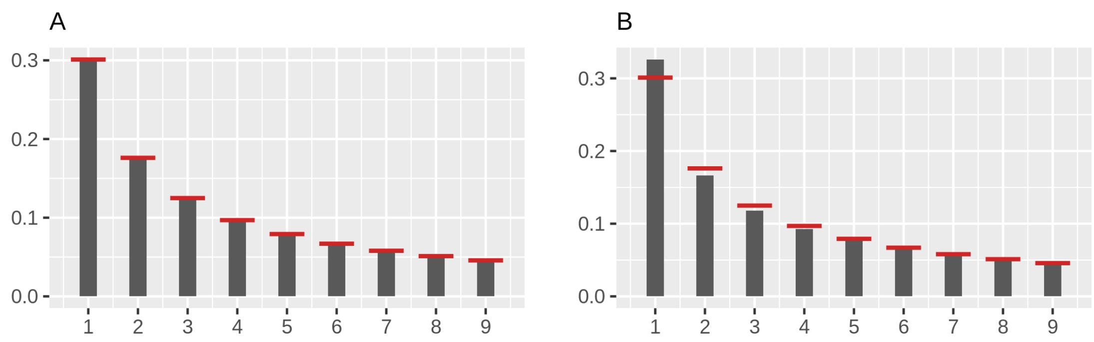

Figure 2.1: Leading digit frequencies of the first 1000 numbers of the Fibonacci sequence (A) and the twitter followers data (B). Red bars: Frequencies predicted by Benford's law.

lotteries, telephone numbers, data sampled from a Gaussian or uniform distribution show that not all numerical data sets follow Benford's law. Durtschi et al. proposed a list of features that indicate if a data set possibly conforms to Benford's law [11]:

1. the data set has to be large
2. the number values have to span a wide range
3. the values have to exhibit a positively skewed distribution
4. the numbers should not be human-assigned

[2]The data is retrieved from `http://testingbenfordslaw.com`, last access 13.12.2019.

Since these characteristics are not supported by mathematical formulas, they are rather vague and should just be seen as rough indicators for identifying Benford distributed data sets.

2.2 Explanations of Benford's law

The first rigorous explanation of Benford's law was proposed by Hill, providing a theoretical proof that data sampled from a mixture of different distributions converges to the Benford distribution [15]. Alternative explanations center around the scale- and base-invariance of the Benford distribution [16, 36], around the spread of the data [12] and around the geometric basis of the law [2][3]. While Hill and Berger wrote a discussion to critically reflect the spread hypothesis, the scale-invariance explanation seems to agree with the others [3, 4]. If the original data is spread out over several orders of magnitude, so would the scaled data. Also the rate of growth would be unchanged by scaling the data.

A different explanation of Benford's law was proposed by Mardia & Jupp [20] and by Smith [42]. Their approaches were based on Poincaré's theorem and Fourier transforms, respectively, but a simpler formulation from Fewster exists [12]. According to them, Benford's law is not an inherent property of the data, but it results from the mathematical transformations applied to get to the significant digit. The more spread the probability density function of the data is on the logarithmic scale and the more uniform it appears there, the better Benford's law will be followed. Pimbley published two articles building on Smith's and Fewster's work [33, 34]. He created a "Benford test" to determine, based on the functional form of the probability density function, whether the resulting distribution of digits would conform to Benford's law [34].

Even though these approaches try to explain from a mathematical perspective how to determine whether the data is supposed to follow Benford's law, the question remains unanswered why natural data sets exhibit these mathematical characteristics. One reason for the occurrence of Benford's law in natural data sets is the fact that many data sets are log-normal distributed, meaning that not the data, but the magnitude of the data follows a normal distribution. If the standard

[3]See Miller's book from 2015 for a more detailed description of these [23].

deviation on the log scale is sufficiently large, i.e. $\sigma > 0.74$, then the mantissas of the logarithms are uniformly distributed on the interval $[0, 1]$, which means that the data is Benford distributed. Another reason is that many data sets are themselves mixtures of multiple distributions or products of different random variables [32], which converge to the Benford distribution according to Hill's limit theorem [15]. Benford's law holds especially for data with underlying growth processes, which are common in nature. Examples are radioactive decay, bacteria population growth or the Fibonacci numbers [23].

2.3 Deriving the general significant-digit law

When inspecting data, humans extract the first digit of a number at first glance, but for implementing Benford's law, the extraction process has to be represented by a mathematical function. The first step is to normalise each number to the interval $[1, 10)$. Let $x \in \mathbb{R}$. The significand function $S(x) : \mathbb{R} \to [1, 10)$ is defined as

$$S(x) = t, \quad t \in [1, 10), \tag{2.1}$$

where t is the unique number such that $|x| = t \cdot 10^s$ for a unique $s \in \mathbb{Z}$. Here s is called the scale of x. $S(x) = 0$ only if $x = 0$. For example, $S(314.15) = 3.1415 \cdot 10^2$, so $t = 3.1415$ and $s = 2$. The scale of $x > 0,\ x \in \mathbb{R}$, is given by $s = \lfloor \log_{10}(x) \rfloor$, where $\lfloor x \rfloor$ is the floor function, e.g. $\lfloor 3.1415 \rfloor = 3$. The goal is to find a representation of $S(x)$ such that we have direct access to all significant digits x is composed of. To achieve this, we first have to rewrite x:

$$\begin{aligned}
& & x &= t \cdot 10^s & & |t = S(x) \\
&\iff & x &= S(x) \cdot 10^s & & | \cdot 10^{-s} \\
&\iff & S(x) &= x \cdot 10^{-s} & & |x = 10^{\log_{10}(x)} \\
&\iff & S(x) &= 10^{\log_{10}(x)} \cdot 10^{-\lfloor \log_{10}(x) \rfloor} & & \\
&\iff & S(x) &= 10^{\log_{10}(x) - \lfloor \log_{10}(x) \rfloor}, & &
\end{aligned}$$

ending up with an equation for $S(x)$ that does only depend on x. Now let $D_m(x) = d_m,\ m \in \mathbb{N}$ be the mth significant digit of x. Recall, that $d_1 \in \{1, 2, ..., 9\}$ for $m =$

1, and $d_m \in \{0, 1, ..., 9\}$ for $m > 1$, since the first significant digit d_1 is by convention never 0. We observe, that

$$\begin{aligned} S(x) &= d_1 + d_2 \cdot 10^{-1} + d_3 \cdot 10^{-2} + ... + d_m \cdot 10^{1-m} \\ &= \sum_{m \in \mathbb{N}} d_m \cdot 10^{1-m} \end{aligned} \tag{2.2}$$

is the decimal expansion of the significand function. We already know, that e.g. $S(12345.67) = 1.234567$. Using Eq. (2.2) we are able to write:

$$S(12345.67) = 1 + 2 \cdot 10^{-1} + 3 \cdot 10^{-2} + 4 \cdot 10^{-3} + 5 \cdot 10^{-4} + 6 \cdot 10^{-5} + 7 \cdot 10^{-6}.$$

The last step is to get access to every single significant digit. Applying the floor function results in:

$$D_m(x) = \lfloor 10^{m-1} S(x) \rfloor - 10 \lfloor 10^{m-2} S(x) \rfloor, \text{ for all } m \in \mathbb{N}. \tag{2.3}$$

For the first significant digit, Eq. (2.3) reduces to $D_1(x) = \lfloor S(x) \rfloor$.

Benford's law and the uniform distribution

A random variable X satisfies Benford's law, if $S(X)$ has a logarithmic distribution:

$$\text{Prob}(S(X) \leq t) = \log_{10}(t), \; t \in [1, 10), \tag{2.4}$$

where $S(X)$ is a continuous random variable. There is some interesting fact we can discover upon substituting $\log_{10} t = s, \; s \in [0, 1)$:

$$\begin{aligned} & \text{Prob}(S(X) \leq t) = \log_{10}(t) && | \log_{10} t = s, t = 10^s \\ \iff & \text{Prob}(S(X) \leq 10^s) = s && | S(X) = 10^{\log_{10}|x| - \lfloor \log_{10}|x| \rfloor} \\ \iff & \text{Prob}(10^{\log_{10}|X| - \lfloor \log_{10}|X| \rfloor} \leq 10^s) = s && | \log_{10} \\ \iff & \text{Prob}(\log_{10}|X| - \lfloor \log_{10}|X| \rfloor \leq s) = s \\ \iff & \text{Prob}(\log_{10} S(x) \leq s) = s, \end{aligned}$$

telling us that $\log_{10} S(x) \sim \mathcal{U}([0, 1])$, i.e. $S(X)$ has a continuous uniform distribution on the interval $[0, 1]$. This means that Benford's law is observable if and only if the logarithm of $S(X)$ has a uniform distribution on $[0,1]$[4].

Benford's law for the first k digits

Benford's law of the first digit can be extended in order to obtain the second, third, and all following digit probabilities or even their joint probabilities. Using Eq. (2.4) allows us to derive the distribution of all the significand digits D_k. Consider the first significand digit D_1 and the event $A = \{D_1 = d_1\}$. A is equivalent to the event $\{d_1 \leq S(X) < d_1 + 1\}$. When we want to calculate the probability of A, we can write:

$$\begin{aligned}
\text{Prob}(\{D_1 = d_1\}) &= \text{Prob}(\{d_1 \leq S(X) < d_1 + 1\}) \\
&= \text{Prob}(S(X) \leq d_1 + 1) - \text{Prob}(S(X) \leq d_1) \quad | \text{ Eq. } (2.4) \\
&= \log_{10}(d_1 + 1) - \log_{10}(d_1) \\
&= \log_{10}\left(\frac{d_1 + 1}{d_1}\right) \\
&= \log_{10}\left(1 + \frac{1}{d_1}\right).
\end{aligned}$$

Note that this is exactly Eq. (1.1). We can even extend Eq. (2.4) further and calculate the joint distribution of D_1 and D_2:

$$\begin{aligned}
\text{Prob}(D_1 = d_1, D_2 = d_2) &= \text{Prob}(d_1 + d_2 \cdot 10^{-1} \leq S(X) < d_1 + (d_2 + 1) \cdot 10^{-1}) \\
&= \log_{10}(10 \cdot d_1 + d_2 + 1) - \log_{10}(10 \cdot d_1 + d_2) \\
&= \log_{10}\left(1 + \frac{1}{10 \cdot d_1 + d_2}\right).
\end{aligned} \tag{2.5}$$

With Eq. (2.5), we can compute the probability of all combinations of the two first digits, e.g.

$$\begin{aligned}
\text{Prob}(D_1 = 4, D_2 = 2) &= \text{Prob}(4 + 2 \cdot 10^{-1} \leq S(X) < 4 + (2 + 1) \cdot 10^{-1}) \\
&= \log_{10}(43) - \log_{10}(42) \\
&= 0.0102.
\end{aligned}$$

[4]The uniform distribution is indeed on $[0,1]$ and not on $[0,1)$, since $S(X)$ is a continuous random variable and therefore $\text{Prob}(S(X) < t) = \text{Prob}(S(X) \leq t)$, because $\text{Prob}(S(X) = t) = 0$.

We can calculate the marginal distribution of D_2 by summing over all possible values of D_1:

$$\text{Prob}(D_2 = d_2) = \sum_{k=1}^{9} \log_{10}\Big(1 + \frac{1}{10k + d_2}\Big) \text{ for all } d_2 = 0, 1, ..., 9. \tag{2.6}$$

The marginal distribution of D_2 already starts to resemble a discrete uniform distribution (Fig. 2.2A). Stating Benford's law for the first k digits, we will see that the marginal distributions for further digits $D_k,\ k > 2$, are nearly indistinguishable from the uniform distribution (Fig. 2.2B). Diaconis showed formally that as k gets larger, the distribution of D_k converges in exponential time to the uniform distribution [10].

Definition For all initial sequences of k significant digits, the joint distribution of these first k digits is:

$$\begin{aligned} \text{Prob}(D_1 = d_1,\ D_2 = d_2, ...,\ D_k = d_k) &= \log_{10}\Big(1 + \frac{1}{10^{k-1}d_1 + 10^{k-2}d_2 + ... + d_k}\Big) \\ &= \log_{10}\bigg(1 + \Big(\sum_{m=1}^{k} 10^{k-m} d_m\Big)^{-1}\bigg), \end{aligned} \tag{2.7}$$

where $d_1 \in \{1, 2, ..., 9\}$ and $d_m \in \{0, 1, ..., 9\}$ for all $m \geq 2$ and $D_2, D_3, ..., D_k$ represent the second, third, ..., kth significant digit. Eq. (2.7) enables us to compute the joint probabilities of the first k significant digits of numbers, e.g.

$$\text{Prob}(D_1 = 3, D_2 = 1, D_3 = 4) = \log_{10}\Big(1 + \frac{1}{10^2 \cdot 3 + 10 \cdot 1 + 4}\Big) = 0.001380.$$

2.4 Characterising Benford's law

2.4.1 Scale-invariance

Recall the definition of scale-invariance. A random variable X is **scale-invariant**, if X and αX have the same distribution, where $\alpha > 0,\ \alpha \in \mathbb{R}$. It can be proven that only the random variable that is zero a.s., i.e. $\mathbb{P}(X = 0) = 1$, can be scale-invariant: $\mathbb{P}(X = \alpha X) = 1$. While no positive random variable X can be scale-invariant, this

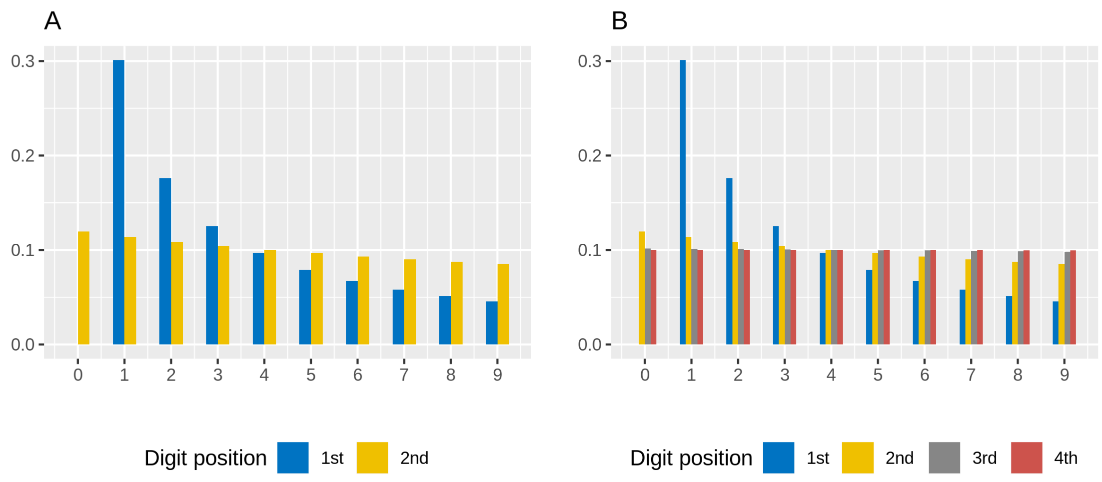

Figure 2.2: Probabilities of the first two (A) and first four significant digits (B), as implied by Benford's law for the first k digits (Eq. (2.7)).

does not have to hold for the distribution of its significant digits, which is what Benford's law is all about. In fact Hill proved, that if and only if X is Benford distributed, its significant digits are scale-invariant [15][5]. In mathematical terms, he showed that if and only if for $\alpha > 0$ and $d \in \{1, 2, ..., 9\}$

$$\mathbb{P}(D_1(X) = d) = \mathbb{P}(D_1(\alpha X) = d) = \log\left(1 + \frac{1}{d}\right), \tag{2.8}$$

then X is Benford. Consider the Fibonacci numbers and scale each of the first 1000 Fibonacci numbers by $\alpha = 2$. If we were to multiply each Fibonacci number by 2, then all numbers with first digits 5, 6, 7, 8, or 9 are converted into numbers starting with 1. So the scaled first digit frequency for 1 should be the sum of the frequencies for the first digit being 5, 6, 7, 8, and 9, which is indeed true as can be seen in Eq. (2.9).

$$\begin{aligned}\text{Prob}(D_1 = 1) &= \text{Prob}(D_1 = 5) + \text{Prob}(D_1 = 6) + \text{Prob}(D_1 = 7) \\ &\quad + \text{Prob}(D_1 = 8) + \text{Prob}(D_1 = 9) \\ &= 0.079 + 0.067 + 0.058 + 0.051 + 0.046 \\ &= 0.301.\end{aligned} \tag{2.9}$$

[5]Not providing mathematical proofs in this section, the interested reader should consult the book from Hill and Berger, which contains all the relevant proofs, written in a strictly mathematical and formal way [5].

Continuing this procedure for the other digits, the scaled Fibonacci numbers are again Benford distributed.

2.4.2 Base-invariance

The most general form of Benford's law can be obtained by treating the base as a variable. Benford's law does not only hold for base 10, but for any base as can be seen in Eq. (2.10) [16]. Here $\log_b$ denotes the base-b logarithm and $D_1^{(b)},\ D_2^{(b)},\ \ldots,\ D_k^{(b)}$ are the first, second, ..., kth significant digits in base b, $b \in \mathbb{N}, b \geq 2$. Thus d_1 is an integer in $\{1,\ 2,\ \ldots,\ b-1\}$ and $d_2,\ \ldots,\ d_k$ are integers in $\{0,\ 1,\ \ldots,\ b-1\}$.

$$\mathrm{Prob}\Big(\big(D_1^{(b)},\ D_2^{(b)}, \ldots,\ D_k^{(b)}\big) = \big(d_1,\ d_2,\ \ldots,\ d_k\big)\Big) = \log_b\left(1 + \left(\sum_{m=1}^{k} b^{k-m} d_m\right)^{-1}\right). \tag{2.10}$$

Note, for $k = 1$ and $b = 2$, Eq. (2.10) reduces to $\mathrm{Prob}\big(D_1^{(2)} = 1\big) = 1$, which is true because the first significant non-zero digit in base 2 is always 1[6].

2.4.3 Power theorem

There is another remarkable property about Benford's law. The classical continuous distributions like the normal and uniform distribution, do not conform to Benford's law (Fig. 2.3). However, if we sample from the standard normal distribution $\mathcal{N}(\mu =$

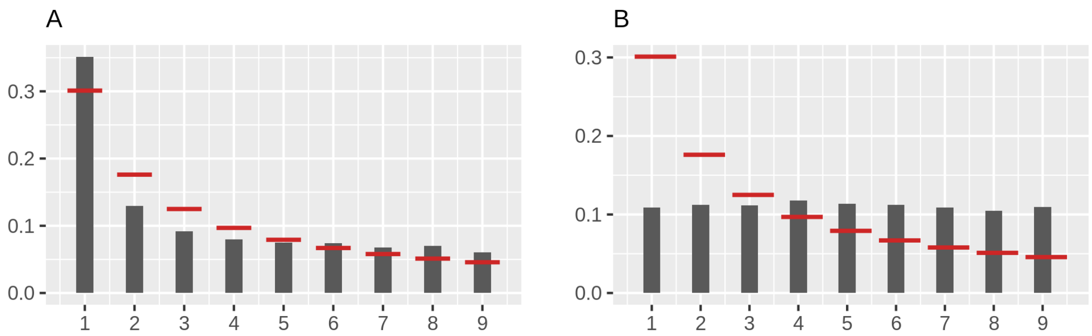

Figure 2.3: Leading digit frequencies of 10000 samples from a standard normal distribution (A) and 10000 samples from a uniform distribution $\mathcal{U}(0, 1)$ (B). Red bars: Frequencies predicted by Benford's law.

[6] Of course when converting the numbers to a different base, the set of first digits and their frequencies will be different. However, if the original data is conform to Benford's law, the logarithmic decay in frequencies will be preserved.

0, $\sigma^2 = 1$) and from the continuous uniform distribution $\mathcal{U}(0,1)$, and raise the samples of the normal distribution to the power of 3 and the samples of the uniform distribution to the power of 35, the resulting first digit distributions are nearly indistinguishable from the Benford distribution (Fig. 2.4).

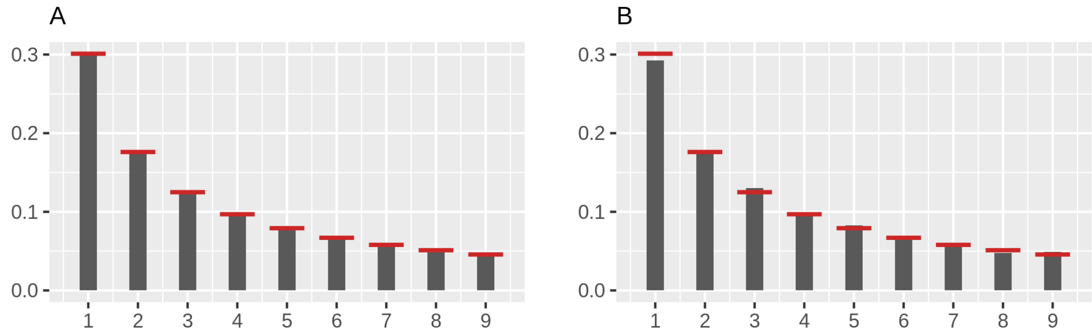

Figure 2.4: Leading digit frequencies of 10000 samples from a standard normal distribution raised to the power of 3 (A) and from a uniform distribution $\mathcal{U}(0,1)$ raised to the power of 35 (B). Red bars: Frequencies predicted by Benford's law.

It can be proven that if X is a continuous random variable with a probability density function, then the significands of the sequence X^n, $n = 1, 2, 3, \ldots$, i.e. $S(X^n)$, converge in distribution to the logarithmic distribution in Eq. (2.4) [5].

Chapter 3

Benford or not Benford

Despite the fact that no finite data set can conform exactly to Benford's law as stated in Eq. (2.7)[1], it is still possible to test whether the data approximately follows Benford's law.

3.1 Methods to test a hypothesis

There are two main paradigms of statistics, called frequentist and Bayesian statistics. Plenty of ink has been spilled on these two approaches, so we will not provide a detailed discussion, but rather keep it to the most relevant information in regard to Benford's law. The difference between frequentist and Bayesian statistics is their respective concept of probability. For the frequentist statistician, a probability is the limiting value of the relative frequency of an event after repeating it many times. For the Bayesian statistician, probabilities represent the degree of belief in a certain event or hypothesis. These distinct conceptions result in different approaches to test a research hypothesis.

3.1.1 The frequentist approach

In the classical approach to hypothesis testing, also called Null Hypothesis Significance Testing (NHST), there is a null hypothesis H_0, in the context of Benford's law stating that the data follows Benford's law and an alternative hypothesis H_1, stating

[1] No finite data set can conform exactly to Benford's law, because the digit frequencies we obtain by inspecting our data will be rational numbers, but the frequencies given by Benford's law are real numbers.

that the data does not follow Benford's law. In NHST, the whole inference process is based on the assumption that the null hypothesis was true. Then via a test statistic, a p-value is calculated, denoting the probability of getting a value from the test statistic that is as extreme or more extreme than the actual obtained value. If this probability is sufficiently small, say smaller than the predefined significance level $\alpha = 0.05$, we decide to reject the null hypothesis. The significance level α denotes the probability that the researcher falsely rejects the true null hypothesis. Note that this whole reasoning is based on the space of possible outcomes generated by the null hypothesis, where the space of possible outcomes depends on the researcher's intention behind the data generating process. Well known frequentist hypothesis tests are the χ^2-test, the Kolmogorov-Smirnov test and the Kuiper test. Not connected to the hypothesis-testing framework are distance-based tests such as the $d*$ distance, which measures the euclidean distance from the first digit frequencies of the data to the Benford predicted frequencies [8][2].

There exists substantial criticism against NHST. The main objection is that p-values do not represent the magnitude of deviation from the null hypothesis [45]. Any deviation can produce a small p-value if the sample size is big enough, rendering it nearly impossible to not reject the null hypothesis [31]. On the other hand, a small sample size can cause large deviations from Benford's law, which are not evaluated as statistically significant. Furthermore, it is impossible to quantify the evidence the p-value provides against the null hypothesis [19]. The p-value is not the probability of the null hypothesis being true, on the contrary, it already assumes the null hypothesis to be true and it only indicates whether the predefined significance level α was exceeded or not. This dichotomous classification of conformance to Benford's law is another disadvantage. Since in practical applications one is gathering evidence for a potential fraud through non-conformance to Benford's law, it would be more efficient to measure the magnitude of deviance and to include prior knowledge about potentially manipulated data.

[2] See Morrow's review for the use of these tests to Benford's law [24].

3.1.2 The Bayesian approach

Contrary to a frequentist statistician, a Bayesian statistician places probabilities in certain parameter values and updates her belief by computing the posterior probability of the parameter value θ, given the observed data $\mathcal{D}$ using Bayes' rule (Eq. (3.1)).

$$\underbrace{\mathbb{P}(\theta|\mathcal{D})}_{\text{posterior}} = \frac{\overbrace{\mathbb{P}(\mathcal{D}|\theta)}^{\text{likelihood}} \cdot \overbrace{\mathbb{P}(\theta)}^{\text{prior}}}{\underbrace{\mathbb{P}(\mathcal{D})}_{\text{marginal}}}. \tag{3.1}$$

There are two ways for a Bayesian statistician to test a hypothesis. One is to calculate whether the parameter value of interest θ_0 is one of the credible values in the posterior distribution, which is commonly known as parameter estimation. The other way is to set up two models, M_0 and M_1, with different prior distributions, where the prior distribution of M_0 places most credibility in the parameter value of interest θ_0 and the prior distribution of M_1 distributes the credibility over many possible parameter values. The most prominent Bayesian model comparison method is the Bayes factor. For the Bayesian models M_0 and M_1, which both intend to explain some data $\mathcal{D}$, the Bayes factor in favour of M_0 is defined as

$$BF_{01} = \frac{\mathbb{P}(\mathcal{D}|M_0)}{\mathbb{P}(\mathcal{D}|M_1)} = \frac{\int \mathbb{P}(\mathcal{D}|\theta_i, M_0)\mathbb{P}(\theta_i|M_0)\mathrm{d}\theta_i}{\int \mathbb{P}(\mathcal{D}|\theta_j, M_1)\mathbb{P}(\theta_j|M_1)\mathrm{d}\theta_j}.$$

Thus the Bayes factor is the ratio of the marginalised prior likelihoods of the data under each model. Note that the Bayes factor compares the models before they have seen the data. It also implicitly accounts for the complexity of a model. Consider two models, where one has more parameters than the other. Since the Bayes factor integrates over all possible parameter values and the more complex model has to distribute its prior probability of the parameters over a broader area, the likelihoods of the complex model will be weighted by a smaller value. This renders it generally more difficult for the complex model to obtain a higher marginal probability.

Integrating over the whole parameter space, the Bayes factor is computationally expensive. Fortunately, it can be computed directly without much effort using the Savage-Dickey method. The Savage-Dickey method permits computing the Bayes factor for nested models, where the nesting model contains all the parameters of the nested model and possibly more. For a properly nested model M_0 under the model M_1, s.t. M_0 assigns the values $\vec{t} = \langle t_1, ..., t_k \rangle$ to the parameters $\theta_1, ..., \theta_k$ of M_1, the Bayes factor in favour of M_0 over M_1 is defined as

$$BF_{01} = \frac{\mathbb{P}(\theta_1 = t_1, ..., \theta_k = t_k | \mathcal{D}, M_1)}{\mathbb{P}(\theta_1 = t_1, ..., \theta_k = t_k | M_1)}.$$

Hence we obtain the Bayes factor with the Savage-Dickey method by dividing the posterior probability by the prior probability at the position of the fixed parameter values $t_1, ..., t_k$ under the nesting model M_1.

3.2 Bayesian model for Benford data

Commonly, in digit analysis using Benford's law the first and second digits are inspected. Therefore, we will build a model for the first and one for the second digit distribution. In a data set consisting of N non-zero numbers, there are N first digits D_1, where D_1 can take one of $k = 9$ possible values, i.e. $d_1 \in \{1, 2, 3, ..., 9\}$. In this data set there are also M second digits D_2, where $M \leq N$ and $d_2 \in \{0, 1, 2, ..., 9\}$, hence the number of categories is $k = 10$. Modelling our data, we assume that the first and second digit frequencies were generated by a nine-variate and by a ten-variate multinomial probability mass function, respectively. In order to carry out Bayesian inferences, we additionally need prior distributions over the parameter space. The Dirichlet distribution, being the conjugate to the multinomial distribution, will be our prior distribution of choice. A conjugate prior is of advantage for Bayesian analysis, since it provides a closed-form expression for the posterior distribution, without the necessity to compute the marginal distribution $\mathbb{P}(\vec{x})$ of the data from Eq. (3.1).

3.2.1 The Multinomial-Dirichlet model

Let $\vec{X} = \langle X_1, ..., X_k \rangle$ be the random vector denoting the number of sampled digits in each category, and $\vec{x} = \langle x_1, ..., x_k \rangle$ be a data sample from this random vector. The probability that an unseen number will be in category i is denoted by θ_i, where $i = 1, ..., k$. $\vec{\theta} = \langle \theta_1, ..., \theta_k \rangle$ is the probability vector over all categories. It has to hold that $\theta_i \geq 0$ and $\sum_{i=1}^{k} \theta_i = 1$, thus the parameter space is $\Theta = [0, 1]^k$.

Let $\vec{X} \sim \mathcal{M}ultinomial(\vec{\theta}, N)$ with probability mass function

$$f_{\vec{\theta},\vec{X}}(x_1, ..., x_k) = \frac{\Gamma(\sum_{i=1}^{k} x_i + 1)}{\prod_{i=1}^{k} \Gamma(x_i + 1)} \prod_{i=1}^{k} \theta_i^{x_i}, \tag{3.2}$$

where $\Gamma(\cdot)$ is the gamma function and all $x_i \geq 0$. In this model, the sample space is $\Omega = \mathbb{N}^k$, the σ-algebra is $\mathcal{E} = \mathcal{P}(\mathbb{N}^k)$ and the probability measure is $\mathbb{P}_{\vec{X}}(x_1, ..., x_k|\vec{\theta}) = f_{\vec{\theta},\vec{X}}(x_1, ..., x_k)$, resulting in the statistical model $(\Omega, \mathcal{E}, \mathbb{P}_{\vec{\theta}})_{\vec{\theta} \in \Theta}$. We choose the prior distribution $\mathbb{P}(\vec{\theta})$ over the parameter space to be the Dirichlet distribution, as it is the conjugate prior to the multinomial distribution. The probability density function of $\vec{\theta} = \langle \theta_1, ..., \theta_k \rangle \sim \mathcal{D}irichlet(\vec{\alpha} = \langle \alpha_1, ..., \alpha_k \rangle)$, with $\forall \alpha_i > 0$, is

$$g_{\vec{\alpha}}(\theta_1, ..., \theta_k) = \frac{1}{B(\vec{\alpha})} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1},$$

where $B(\vec{\alpha}) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)}$ and $\Gamma(\cdot)$ is the gamma function. Using the conjugate prior for the multinomial distribution means that the posterior distribution $\mathbb{P}(\vec{\theta}|\vec{x})$ will again be a Dirichlet distribution, where the values of $\vec{\alpha}$ are updated by adding the values of $\vec{x}$ to it:

$$\vec{\theta} \mid \langle x_1, ..., x_k \rangle \sim \mathcal{D}irichlet(\alpha_1 + x_1, ..., \alpha_k + x_k).$$

Sketching the proof of the Dirichlet distribution being the conjugate to the multinomial distribution, we consider the probability mass function of the multinomial distribution $f_{\vec{\theta},\vec{X}}(x_1, ..., x_k)$ and the probability density function of the Dirichlet distribution $g_{\vec{\alpha}}(\theta_1, ..., \theta_k)$. Upon neglecting the normalisation factors that do not depend on θ we get

$$f_{\vec{\theta},\vec{X}}(x_1, ..., x_k) \propto \prod_{i=1}^{k} \theta_i^{x_i}, \quad g_{\vec{\alpha}}(\theta_1, ..., \theta_k) \propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1},$$

resulting in

$$\begin{aligned} f_{\vec{\theta},\vec{X}}(x_1, ..., x_k) \cdot g_{\vec{\alpha}}(\theta_1, ..., \theta_k) &\propto \prod_{i=1}^{k} \theta_i^{x_i} \cdot \prod_{i=1}^{k} \theta_i^{\alpha_i - 1} \\ &= \prod_{i=1}^{k} \theta_i^{x_i + \alpha_i - 1} \\ &\propto g_{(\langle \alpha_1 + x_1, ..., \alpha_k + x_k \rangle)}(\theta_1, ..., \theta_k). \end{aligned}$$

Summing up, in the Multinomial-Dirichlet model there is a multinomial distributed random vector $\vec{X}$ with the parameters $N = \sum_{i=1}^{k} X_i$ and $\vec{\theta}$, where $\vec{\theta}$ is Dirichlet distributed with parameter $\vec{\alpha}$ (Fig. 3.1).

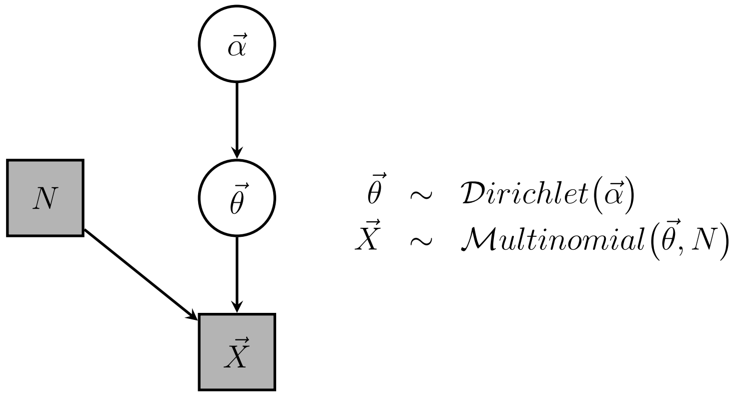

Figure 3.1: Visualisation of the Multinomial-Dirichlet model, following the graphical conventions from Lee and Wagenmakers [18].

3.2.2 Prior distributions

Opponents of the Bayesian approach criticise the need to specify a prior distribution over the parameter space for bringing a subjective component into the hypothesis testing procedure. However, this is of advantage for us, since we can choose the Dirichlet distribution to express different degrees of uncertainty by adjusting the values for $\vec{\alpha}$.

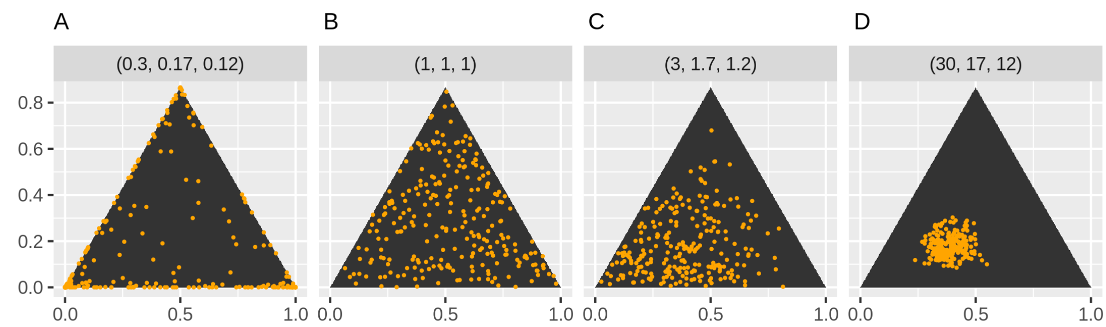

Figure 3.2: Probability density of a Dirichlet distribution with different $\vec{\alpha}$, expressing a varying degree of uncertainty about the parameter space.

If we have no prior information about the data, we can distribute the probability density uniformly over every possible parameter combination by setting $\vec{\alpha} = \langle \alpha_1 = \alpha_2 = ... = \alpha_k = 1 \rangle$, yielding an "uninformed prior" (Fig. 3.2B). The most intuitive prior in regard to digit analysis with Benford's law is certainly an "informed prior" that already places probabilities in the Benford predicted frequencies. For the first digit distribution, we set $\vec{\alpha}$ to the Benford predicted frequencies for the first digit and multiply it by some scalar $\beta \in \mathbb{R}$: $\vec{\alpha} = \langle \alpha_1 = \beta \cdot 0.301, \alpha_2 = \beta \cdot 0.176, ..., \alpha_9 = \beta \cdot 0.046 \rangle$, where β is chosen such that $\exists \alpha_i > 1$ (Fig. 3.2C). The higher β, the higher is our prior belief in the data to be Benford (Fig. 3.2D). If we test for the second or some other digit position, the values of $\vec{\alpha}$ will of course correspond to those postulated by Benford's law for this respective digit position.

Another established prior for Bayesian data analysis is the "objective prior". In case we had a prior belief about the data not being Benford distributed, we would set $\vec{\alpha} = \langle \alpha_1 = 0.301, \alpha_2 = 0.176, ..., \alpha_9 = 0.046 \rangle$, and multiply $\vec{\alpha}$ by some scalar $\beta \in \mathbb{R}$, where β is chosen such that $\forall \alpha_i < 1$, which means that the parameters are getting pushed towards extreme values (Fig. 3.2A). For our purpose the objective prior is however not useful, since we want to place initial belief around the parameter values representing the null hypothesis that the data is Benford distributed. Therefore, we would rather choose an informed prior with a relatively high value of β, to emphasise the null hypothesis that the data is Benford distributed.

Tab. 3.1 serves as a point of reference for the effects different values of β have to the prior distribution for the first significant digit.

prior	objective	weakly informed	informed	strongly informed
$\beta \in$	(0, 3.3]	(3.3, 22]	(22, 150]	(150, ∞)

Table 3.1: The types of priors to which different values of β correspond, applied to the first digit distribution.

3.2.3 Bayes factors via the Savage-Dickey method

The Bayes factor enables the Bayesian statistician to gather continuous evidence in favour of or against the null hypothesis that a data set follows the Benford distribution. The null hypothesis is represented by the fixed parameter vector

$$\vec{t} = \langle 0.301, 0.176, 0.125, 0.097, 0.079, 0.067, 0.058, 0.051, 0.046 \rangle$$

for the first significant digit and

$$\vec{t} = \langle 0.12, 0.114, 0.109, 0.104, 0.10, 0.097, 0.093, 0.090, 0.088, 0.085 \rangle$$

for the second significant digit. Using the Multinomial-Dirichlet model to describe the data, we have to set up two nested models in order to compute the Bayes factor with the Savage-Dickey method. The nested model M_0 is composed of the likelihood

$$\mathbb{P}_{M_0}(\vec{x}|\theta_1 = t_1, ..., \theta_k = t_k) = f_{\vec{t},\vec{X}}(x_1, ..., x_k) = \frac{\Gamma(\sum_{i=1}^k x_i + 1)}{\prod_{i=1}^k \Gamma(x_i + 1)} \prod_{i=1}^k t_i^{x_i},$$

and the prior distribution

$$\mathbb{P}_{M_0}(\theta_1 = t_1, ..., \theta_k = t_k) = g_{\vec{\alpha}}(\theta_1 = t_1, ..., \theta_k = t_k) = \frac{1}{B(\vec{\alpha})} \cdot \prod_{i=1}^k t_i^{\alpha_i - 1}.$$

The nesting model M_1 is composed of the likelihood

$$\mathbb{P}_{M_1}(\vec{x}|\theta_1, ..., \theta_k) = f_{\vec{\theta},\vec{X}}(x_1, ..., x_k) = \frac{\Gamma(\sum_{i=1}^k x_i + 1)}{\prod_{i=1}^k \Gamma(x_i + 1)} \prod_{i=1}^k \theta_i^{x_i},$$

and the prior distribution

$$\mathbb{P}_{M_1}(\theta_1, ..., \theta_k) = g_{\vec{\alpha}}(\theta_1, ..., \theta_k) = \frac{1}{B(\vec{\alpha})} \cdot \prod_{i=1}^k \theta_i^{\alpha_i - 1},$$

yielding the posterior distribution

$$\mathbb{P}_{M_1}(\theta_1, ..., \theta_k|\vec{x}) = g_{(\vec{\alpha}+\vec{x})}(\theta_1, ..., \theta_k) = \frac{1}{B(\vec{\alpha} + \vec{x})} \cdot \prod_{i=1}^k \theta_i^{\alpha_i + x_i - 1},$$

since the Dirichlet distribution is conjugate to the multinomial distribution. Inserting $\vec{t} = \vec{\theta}$ in the prior and the posterior distribution of M_1 yields the Bayes factor in favour of M_0 over M_1:

$$\begin{aligned} BF_{01} &= \frac{\mathbb{P}(\theta_1 = t_1, ..., \theta_n = t_k|\mathcal{D}, M_1)}{\mathbb{P}(\theta_1 = t_1, ..., \theta_k = t_k|M_1)} \\ &= \frac{g_{(\vec{\alpha}+\vec{x})}(\theta_1 = t_1, ..., \theta_k = t_k)}{g_{\vec{\alpha})}(\theta_1 = t_1, ..., \theta_k = t_k} \\ &= \frac{\frac{1}{B(\vec{\alpha}+\vec{x})} \cdot \prod_{i=1}^k t_i^{\alpha_i + x_i - 1}}{\frac{1}{B(\vec{\alpha})} \cdot \prod_{i=1}^k t_i^{\alpha_i - 1}} \\ &= \frac{B(\vec{\alpha})}{B(\vec{\alpha} + \vec{x})} \cdot \prod_{i=1}^k t_i^{x_i}. \end{aligned} \tag{3.3}$$

Chapter 4

Round digits

4.1 Human generated numbers

If a numerical data set is supposed to conform to Benford's law, then a manipulation of the data will yield deviations from the Benford predicted frequencies. There are several reasons that hinder humans from generating Benford distributed data. One practical reason is that humans are unaware of the existence of the underlying Benford distribution [8]. Nevertheless, it can be shown that even if people know the data generating process and the underlying distribution, they are unsuccessful in sampling data conforming to it [6]. Furthermore, humans try to avoid writing the same digit multiple times in a row, but from a probabilistic perspective, Benford distributed data has to exhibit this characteristic [38]. Instead, people tend to repeat the same sequences within several different numbers [23], or the whole number [28]. For instance, they would repeatedly use the number combination 27, 67 and 87, resulting in the generated numbers 127679.87, 987.27, 35467.27. Another reason is that humans subconsciously favour certain numbers and sequences over others [11].

There exists the hypothesis that when people generate numbers, they are biased by the numbers appearing in their everyday life [14]. As a consequence, people would be great Benford number generators if in their lives they were only exposed to Benford distributed data. Testing whether Benford's law is indeed that omnipresent in our lives, Hill conducted an experiment where he let undergraduate students write down random numbers composed of six digits [14]. His results showed that the human generated numbers indicate some conformance to Benford's law, but in general they deviate from it.

Not Benford's law governs the appearance of numbers in our life, but two features immanent in each number: the magnitude [9] and the roundness [17]. The number 2 is more frequently used than the number 28, since 2 is smaller than 28. Nevertheless, the number 28 is not more frequent than the number 100, because it is less round than 100. Roundness is defined as the frequency of the number in an approximate context. In order to understand the way people generate numbers, we will further investigate the roundness property.

4.1.1 Some numbers are rounder than other

In Fig. 4.1 there are digits ranging from 2 to 1000 plotted against their occurrence. A logarithmic decay can be observed, resembling the logarithmic decay of Benford's law. However, there are peak numbers with a much higher occurrence compared

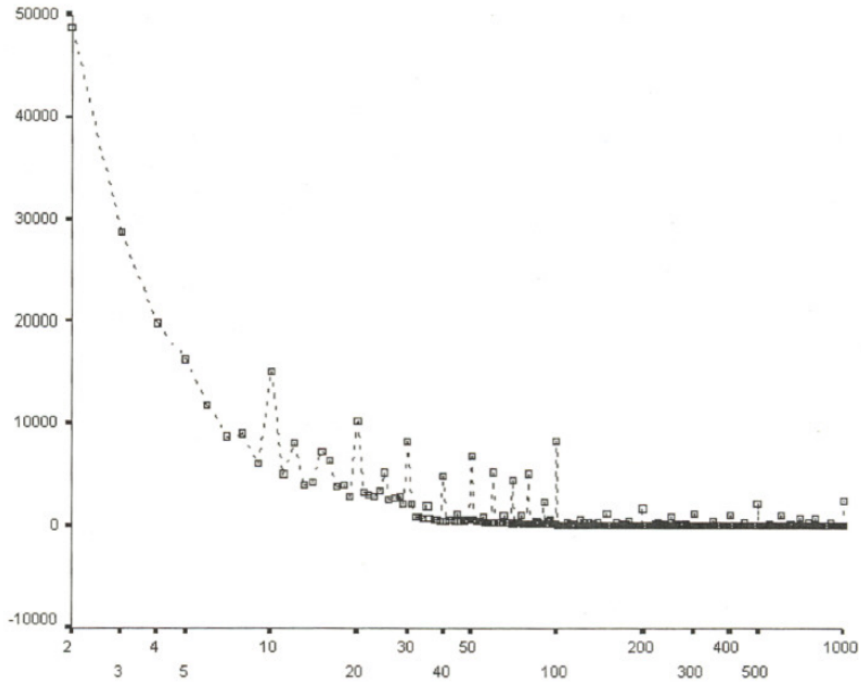

Figure 4.1: Number frequencies in the 27-mln-corpus of the Institute for Dutch Lexicology in Leiden, plotted on a logarithmic scale. Retrieved from Jansen's and Pollmann's article "On Round Numbers: Pragmatic Aspects of Numerical Expressions" [17].

to their neighbours. These peak numbers are for example the numbers 10, 12, 15, 20, 50, 100, and 1000, exactly those numbers that we consider being rounder than others. For cultures with the decimal number system, the degree of roundness of a number is related to its numerical properties. We say that a number x is round, if

$$x \in \{1, ..., 9\} \quad \vee \quad \exists r \in \{2,\ 2.5,\ 5,\ 10\} : \ x \mod r = 0, \tag{4.1}$$

where $x \in \mathbb{N}$. If Eq. (4.1) holds for more than one r, x seems even rounder to humans. Considering the numbers 14, 25, 100 and 101, one can observe that 14 can be divided by 2 without leaving a remainder, 25 by 2.5 and 5, 100 by 2, 2.5, 5 and 10. Thus 100 is rounder than 25 and 25 is rounder than 14. 101 leaves for all r a remainder, so it is not round at all.

Previously we said that roundness refers to numbers being used in an approximate context. It can be shown that roundness and magnitude determine the frequency of occurrence of a number irrespective of its context [17]. This implies that round numbers are ubiquitous in a human's life and therefore favoured when generating numbers manually. Informally confirming this hypothesis, one can think about the numbers on banknotes, analog watches, or round birthdays and anniversaries. To formally confirm the hypothesis that when forging numbers humans favour round numbers as significant digits over others, we conducted an experiment.

4.2 Exploratory behavioral experiment

If the hypothesis that round numbers are favoured by humans was true, we predicted that round digits will be chosen more frequently than others when humans have to forge numbers. The experiment is exploratory and is supposed to indicate which digits exactly are favoured[1].

Participants 50 participants were recruited from Prolific and each was presented with nine items, resulting in 450 responses. In all experiments, only native English speaking participants were allowed who have at least a 75% work approval rating. The experiment took on average 4.2 minutes and participants were compensated £0.60.

Procedure In the experiment, the participants had to imagine being the head of a hospital and being responsible for the financial matters. For the child

[1] The experiment was designed with _magpie and can be found at `https://github.com/Nina-Mainusch/forge-numbers`.

cancer treatment section they have a budget of $1,000,000 per year. There are two conditions, one in which they have already spent too much money and one where they have not spent enough, with the consequence of a shortened budget for the next year. They have no choice but to revise some of the previous expenses downwards or upwards respectively, in order to not lose money for the next year's budget. Then each participant is presented with nine different expenses one at a time and they have to revise them, entering the new amount of money in a text box (Fig. 4.2).

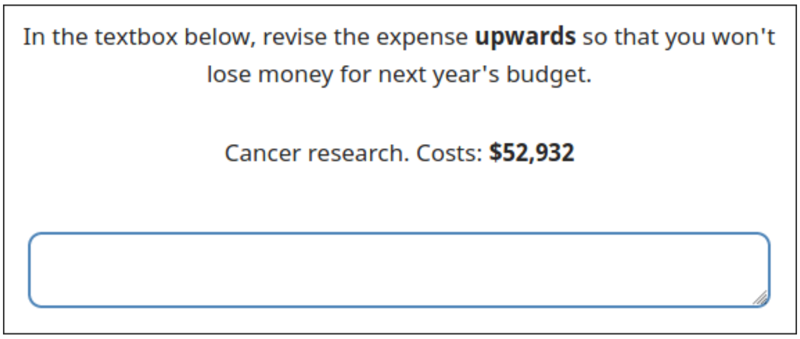

Figure 4.2: Exemplary experimental trial of the "upwards" condition.

Materials The cancer hospital situation was chosen for the participants to not feel bad about forging numbers. The nine items each participant saw were uniformly sampled numbers, ranging from $1,000$ to $999,999$, to not unconsciously bias the participants by certain digits (Tab. 4.1).

1014	4422	7563	17674	34719	50115	66982	80704	106558
2114	6086	8522	25207	40981	55411	72300	87607	512825
3689	6740	13424	31420	47361	64292	77984	94471	844025

Table 4.1: 27 of the 441 items used in the experiment.

Results All participants manipulated their items in the direction as indicated in their task description, but one participant was excluded for self-reporting he did not understand the task, leaving a remainder of 49 participants and 441 responses for the analysis.

To get an overview, we first inspect the digit distribution of every digit present in the data (Fig. 4.3A). Except for the digit 0, all digits indeed center around their predicted frequencies. Digit 0 is less frequent, since the first significant digit is per definition never 0, but only the subsequent digits can be 0. The digit

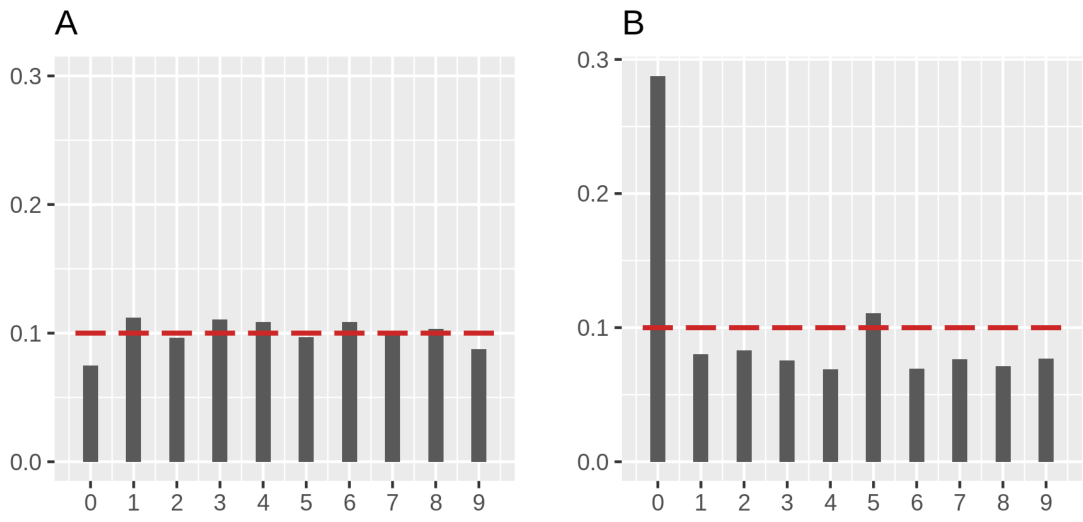

Figure 4.3: Frequencies for all digits of the items (A) and the responses (B). Red bars: uniform probabilities that generated the item's digits.

distribution of the responses on the other hand clearly indicates a human's preference for the digits 0 and 5 (Fig. 4.3B). Closer investigation of the distribution for the first (Fig. 4.4B) and second (Fig. 4.4D) significant digit of the responses revealed that humans prefer the digit 1 as the first significant digit and the digits 0 and 5 as second significant digits. It seems plausible to hypothesise that humans have preferences for the first and second digit when forging numbers, resulting in the alternative hypotheses:

$H_{1.1}$: humans have preferences when forging the first significant digit,

$H_{1.2}$: humans have preferences when forging the second significant digit.

We test these hypotheses with the χ^2-test and by computing Bayes factors using the Bayesian model for analysing digit distributions from chapter 3.

4.2.1 Hypothesis testing with the χ^2-test

Testing $H_{1.1}$ We conducted a χ^2-test assuming the significance level $\alpha = 0.05$, and the null hypothesis that humans have no preferences when forging the first digit, represented by the vector of expected probabilities $\vec{p}_1 = \langle \frac{1}{9}, ..., \frac{1}{9} \rangle$. The observed test result was significant ($\chi^2 = 18.082$, $df = 8$, p-value $= 0.02062$). Based on this, the null hypothesis can be rejected and we conclude that humans have preferences when forging the first digit.

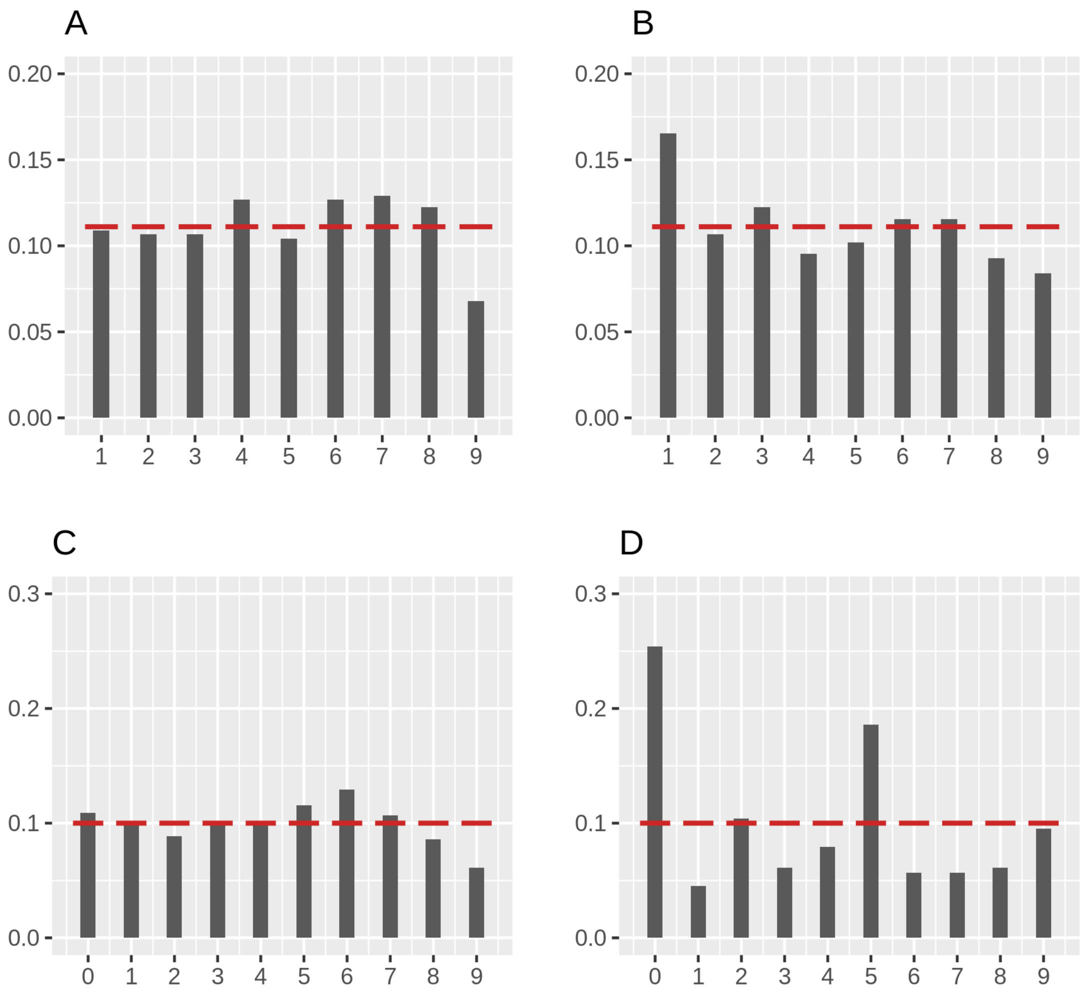

Figure 4.4: First digit frequencies of the items (A) and the responses (B). Second digit frequencies of the items (C) and the responses (D). Red bars: uniform probabilities that generated the item's digits.

Testing $H_{1.2}$ We conducted a χ^2-test assuming the significance level $\alpha = 0.05$, and the null hypothesis that humans have no preferences when forging the second digit, represented by the vector of expected probabilities $\vec{p}_2 = \langle \frac{1}{10}, ..., \frac{1}{10} \rangle$. The observed test result was significant ($\chi^2 = 182.15$, $df = 9$, p-value $< 2.2e^{-16}$). We reject the null hypothesis and conclude that humans have preferences when forging the second digit.

4.2.2 Hypothesis testing with the Bayes factor

To compute the Bayes factor with the Savage-Dickey method in favour of the null hypothesis, represented by M_0, over the alternative hypothesis, represented by M_1, we define for each hypothesis these two nested models and various prior distributions.

Testing $H_{1.1}$ M_0 assumes $\vec{X} \sim \mathcal{Multinomial}(\vec{\theta}, N)$, where $\vec{\theta} = \langle \theta_1 = \frac{1}{9}, ..., \theta_9 = \frac{1}{9} \rangle$ and $N = 441$. $\vec{x} = \langle 73, 47, 54, 42, 45, 51, 51, 41, 37 \rangle$ is the data from the experiment. The nesting model M_1 assumes $\vec{X} \sim \mathcal{Multinomial}(\vec{\theta}, N)$ and different priors for $\vec{\theta}$:

(A) the uninformed prior $\vec{\theta} \sim \mathcal{Dirichlet}(\vec{\alpha} = \langle \alpha_1 = 1, ..., \alpha_9 = 1 \rangle)$

(B) the informed prior $\vec{\theta} \sim \mathcal{Dirichlet}(\vec{\alpha} = \langle \alpha_1 = \beta \cdot \frac{1}{9}, ..., \alpha_9 = \beta \cdot \frac{1}{9} \rangle)$ with

(B.1) $\beta = 20$,
(B.2) $\beta = 100$,
(B.3) $\beta = 500$.

For the prior (A), inserting these values in Eq. (3.3) yields the Bayes factor BF_{01} in favour of M_0 over M_1:

$$\begin{aligned}
\mathrm{BF}_{01} &= \frac{\mathbb{P}(\theta_1 = \frac{1}{9}, ..., \theta_9 = \frac{1}{9} | \vec{x}, M_1)}{\mathbb{P}(\theta_1 = \frac{1}{9}, ..., \theta_9 = \frac{1}{9} | M_1)} \\
&= \frac{\frac{1}{B(\vec{\alpha}+\vec{x})} \cdot \prod_{i=1}^{9} \left(\frac{1}{9}\right)^{\alpha_i + x_i - 1}}{\frac{1}{B(\vec{\alpha})} \cdot \prod_{i=1}^{9} \left(\frac{1}{9}\right)^{1-1}} \\
&= \frac{\frac{1}{B(\langle 74, 48, ..., 38 \rangle)} \cdot \left(\frac{1}{9}\right)^{441}}{\Gamma(9)} \\
&= \frac{421437686}{40320} = 10452.32,
\end{aligned}$$

which is overwhelming evidence in favour of the null hypothesis that humans have no preferences when forging the first significant digit. Performing the exact same calculations with the priors of (B) yields that the more certain we are that the digits are uniformly distributed, the less probable $H_{0.1}$ gets until we found even evidence with (B.3) in favour of the alternative hypothesis $H_{1.1}$ (Tab. 4.2).

Priors	(A)	(B.1)	(B.2)	(B.3)
BF_{01}	10452	376	2.37	2.56^{-1}

Table 4.2: Bayes factor in favour of $H_{0.1}$ for different prior distributions.

Testing $H_{1.2}$ The only difference to testing $H_{1.1}$ is that we have ten categories for the second digit, so M_0 assumes $\vec{\theta} = \langle \theta_1 = \frac{1}{10}, ..., \theta_{10} = \frac{1}{10} \rangle$. The data from the experiment is $\vec{x} = \langle 112, 20, 46, 27, 35, 82, 25, 25, 27, 42 \rangle$ and $N = 441$. For M_1, the α vectors for the uninformed prior and the informed priors are:

(A) the uninformed prior $\vec{\theta} \sim \mathcal{D}irichlet(\vec{\alpha} = \langle \alpha_1 = 1, ..., \alpha_{10} = 1 \rangle)$

(B) the informed prior $\vec{\theta} \sim \mathcal{D}irichlet(\vec{\alpha} = \langle \alpha_1 = \beta \cdot \frac{1}{10}, ..., \alpha_{10} = \beta \cdot \frac{1}{10} \rangle)$ with

(B.1) $\beta = 20$,
(B.2) $\beta = 100$,
(B.3) $\beta = 500$.

The resulting Bayes factor for the uninformed prior (A) is

$$\begin{aligned} \mathrm{BF}_{01} &= \frac{\mathbb{P}(\theta_1 = \frac{1}{10}, ..., \theta_{10} = \frac{1}{10} | \vec{x}, M_1)}{\mathbb{P}(\theta_1 = \frac{1}{10}, ..., \theta_{10} = \frac{1}{10} | M_1)} \\ &= \frac{3.330762e^{-20}}{362880} = 9.178687e^{-26}, \end{aligned}$$

which is overwhelming evidence in favour of the alternative hypothesis that humans have preferences when forging the second significant digit. Applying the priors from (B), we find again convincing evidence that the data does not conform to the uniform distribution, even though we had with (B.3) a strong prior belief that the digits are uniformly distributed (Tab. 4.3).

Priors	(A)	(B.1)	(B.2)	(B.3)
BF_{01}	$9.18e^{-26}$	$1.4e^{-26}$	$3.8e^{-25}$	$3.5e^{-16}$

Table 4.3: Bayes factor in favour of $H_{0.2}$ for different prior distributions.

4.2.3 Discussion of the results

Comparing the results from the Bayes factor calculations and the p-values from the χ^2-test, we observe that they are inconsistent for $H_{1.1}$. The uninformed and weakly informed priors plead in favour of the null hypothesis, but the informed and strongly informed priors indicate to withhold judgement. The p-value can not decide to withhold judgement and since it is below the significance level α, we are advised to reject the null hypothesis.

When testing $H_{1.2}$, the two approaches agree, where the p-value suggests to reject the null hypothesis that there is no preference for the second significant digit. The Bayes factor even provides evidence in favour of the alternative hypothesis that there is a preference for the second significant digit.

Chapter 5

Testing the Bayesian model

5.1 Simulation study

To analyse the Bayes factors calculated with the Savage-Dickey method in the Multinomial-Dirichlet model, we perform a simulation study. The α-error rate of the χ^2-test results and the Bayes factors for genuine Benford first digit data and three different sample sizes ($N = 30$, $N = 100$, $N = 1000$) is inspected, as well as the proportion of significant test results for distorted Benford data. The distortion is created by replacing 10, 20 or 40% of the genuine Benford data with the digits 1 and 5. These digits were chosen based on the experimental findings in chapter 4. The significance level α for the χ^2-test is set to $\alpha = 0.05$ and a Bayes factor in favour of the null hypothesis is interpreted as significant evidence for the alternative hypothesis if it is smaller than $6^{-1} = 0.1667$. Each setting and each test were repeated 10000 times in order to obtain reliable results.

Results for genuine Benford data The χ^2-test results show that for 0% distorted Benford data, the α-error rate of 0.05 is approached by all sample sizes, unsurprisingly corresponding to the predefined significance level $\alpha = 0.05$. The Bayes factors for 0% distorted Benford data contrarily exhibit an α-error of 0 for any prior and a sample size of 30 and 1000, thus they are never in favour of the alternative hypothesis. Only for a sample size of 100 and an informed prior with $\beta = 100$ the Bayes factor wrongly finds evidence for the alternative hypothesis in 1% of the tests (Fig. 5.1, column (A)).

Results for 10% distorted Benford data For 10% distorted Benford data, the χ^2-test rejects the null hypothesis in 98% of the cases for a sample size of

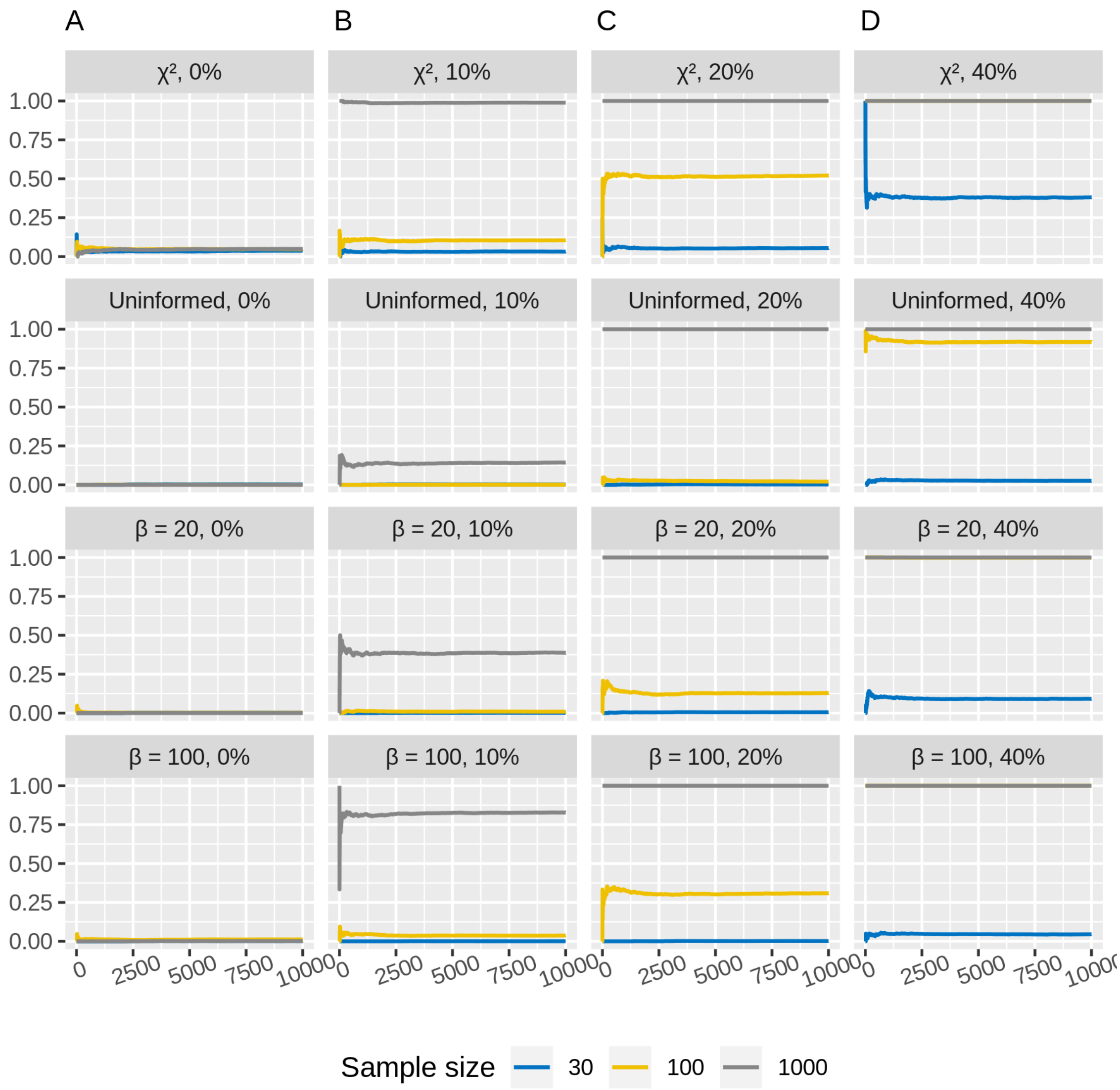

Figure 5.1: Proportion of significant test results for 10000 repetitions and Benford first digit data with 0%, 10%, 20% and 40% distortion. First row: χ^2-test results. Second row: Bayes factor results for an uninformed prior. Third row: Bayes factor results for a weakly informed prior with $\beta = 20$. Fourth row: Bayes factor results for an informed prior with $\beta = 100$.

1000, in 12% of the cases for a sample size of 100 and in 3% of the cases for a sample size of 30. The stronger our prior belief in the data to be Benford distributed, i.e. the larger β is, the more severely does the Bayes factor penalize the 10% distortion in the Benford data for a sample size of 1000. For smaller sample sizes, the alternative hypothesis is barely favoured by any prior (Fig. 5.1, column (B)).

Results for 20% distorted Benford data For 20% distorted Benford data, the χ^2-test always rejects the null hypothesis for a sample size of 1000 and in 50% of the cases for a sample size of 100. The Bayes factor is in favour of the alternative hypothesis for any prior and a sample size of 1000, for a sample size of 100 it gradually is in favour of the alternative hypothesis the stronger our prior belief in the data to be Benford distributed. For a sample size of 30, we cannot find evidence against the null hypothesis neither with the χ^2-test nor with the Bayes factor (Fig. 5.1, column (C)).

Results for 40% distorted Benford data The χ^2-test always rejects the null hypothesis that the 40% distorted data conforms to Benford's law for a sample size of 1000 and 100 and in 37% of the cases for a sample size of 30. The Bayes factor is in favour of the alternative hypothesis for a sample size of 100 and 1000 and all prior distribution specifications, except for the uninformed prior, where it favours the alternative hypothesis in 90% of the cases for a sample size of 100. For a sample size of 30, we can barely find evidence against the null hypothesis for any prior (Fig. 5.1, column (D)).

5.1.1 Discussion of the results

Due to the intrinsic logic of the NHST, we falsely reject the null hypothesis in 5% of the cases for 0% distorted Benford data, which does not occur when testing the hypothesis with the Bayes factor. We experimentally confirmed the claim against the NHST approach that already a minuscule distortion in the data can produce a small p-value if the sample size is big enough. The Bayes factor requires more distortion in the data than the χ^2-test or a strong prior belief in the data to be Benford distributed in order to reject the null hypothesis. However, if we have enough data, the prior distributions do not affect the results of the Bayes factor. Unlike the χ^2-test results, which can only be interpreted as evidence against the null hypothesis, the Bayes factor can provide evidence in both directions, for and against the null hypothesis.

Chapter 6

Conclusion

In the beginning we stated Benford's law and showcased its most important properties. Modelling Benford distributed data, we designed a Bayesian model composed of the multinomial and Dirichlet distribution and calculated a Bayes factor with the Savage-Dickey method to test the conformance of a data set to Benford's law. The Dirichlet distribution being conjugate to the multinomial distribution allows to compute the posterior distribution in one simple arithmetic step. We focused on analysing the first digit distribution, although the model works for any digit position. Since Benford's law is mainly applied to fraud detection, we conducted an experiment to try to understand how humans forge numbers. The more we understand about the data generating process, the better we can model and analyse the potentially fraudulent data. Our findings showed that for humans some numbers feel rounder than others and are more frequently chosen when forging numbers.

To compare the frequentist and the Bayesian approach to hypothesis testing, represented by the χ^2-test and the Bayes factor respectively, we conducted a simulation study for gradually distorted Benford data. Opponents of the Bayesian approach complain that the need to specify prior distributions brings an unfavourable subjective dimension into the Bayesian data analysis. However, the frequentist approach is far from objective itself, for requiring to specify an arbitrary significance level α and for basing its decisions on an imaginary cloud of possible outcomes. The frequentist approach assumes perfect conformance to Benford's law, but a finite data set will never conform perfectly to Benford's law and therefore the usefulness of p-values in the context of fraud detection with Benford's law is reduced. Being sensitive to small distortions in the data, the frequentist approach is additionally prone to produce more and more false-positive results the larger the sample size. The dichotomous decision making in NHST is another disadvantage of the frequentist approach.

The Bayes factor calculated in the Multinomial-Dirichlet model with the Savage-Dickey method on the other hand is a measure of evidence quantifying the magnitude of deviation from the data to Benford's law. It is straightforward to interpret and provides coherent results for all sample sizes. By incorporating prior knowledge about the data distribution, the researcher is encouraged to think about the specific situation which is to be analysed, and to explicitly justify her choices for the prior distribution. The Bayes factor has shown to rather support the null hypothesis than to reject it, compared to the χ^2-test results.

In conclusion, the Bayesian model we designed is a powerful tool to detect deviations from Benford's law, surpassing the performance of the χ^2-test. However, be reminded that we can only apply Benford's law to fraud detection if the underlying data set is supposed to follow Benford's law. Imagine the tax authorities collecting evidence about an accounting fraud by investigating a deviation from a law the accounting data does not even obey due to. The identification of a suspicious data set always requires further investigation and a detected deviation from Benford's law cannot be used as an accusation on its own. For later studies we want to further investigate the effects of the prior distributions to the data analysis, and especially to find optimal values for the parameter β.

Bibliography

[1] Acebo, E. Del, Sbert, M. Benford's Law for Natural and Synthetic Images. In *Computational Aesthetics in Graphics, Visualization and Imaging*. The Eurographics Association, 2005.

[2] Benford, F. The law of anomalous numbers. *Procc. Amer. Philosophical Soc.*, 78:551–572, 1938.

[3] Berger, A. Large Spread Does Not Imply Benford's Law. Preprint, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada, 2010.

[4] Berger, A., Hill, T. P. Benford's law strikes back: no simple explanation in sight for mathematical gem. *Math. Intelligencer*, 33(1):85–91, 2011.

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

[6] Camerer, C. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.

[7] Chiverton, J., Wells, K. *PV Modeling of Medical Imaging Systems*, chapter 18, pages 319–337. Princeton University Press, Princeton, NJ, 2015.

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

[9] Dehaene, S., Mehler, J. Cross-linguistic regularities in the frequency of number words. *Cognition*, 43:1–29, 1992.

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

[11] Durtschi, C., Hillison, W., Pacini, C. The effective use of Benford's Law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, 5:17–34, 2004.

[12] Fewster, R. M. A Simple Explanation of Benford's Law. *The American Statistician*, 63. 1:26–32, 2009.

[13] Geyer, A., Martí, J. Applying Benford's law to volcanology. *Geology*, 40(4):327, 2012.

[14] Hill, T. P. Random-number guessing and the first digit phenomenon. *Psychological Reports*, 62(3):967–971, 1988.

[15] Hill, T. P. A statistical derivation of the significant-digit law. *Statist. Sci.*, 10.4:354–363, 1995.

[16] Hill, T. P. Base-invariance implies Benford's law. *Proc. Amer. Math. Soc.*, 10(4):887–895, 1995.

[17] Jansen, C. J. M., Pollmann, M. M. W. On Round Numbers: Pragmatic Aspects of Numerical Expressions. *Journal of Quantitative Linguistics*, 8:3:187–201, 2001.

[18] Lee, M. D., Wagenmakers, E. *Bayesian cognitive modeling: A practical course.* Cambridge university press, 2014.

[19] Lin, R. and Yin, G. Bayes factor and posterior probability: complementary statistical evidence to p-value. *Contemporary clinical trials*, 44:33–35, 2015.

[20] Mardia, K. V., Jupp, P. E. *Directional Statistics.* Chichester: Wiley, Example 4.1, 2000.

[21] Mebane, W. R. Jr. Comment on "Benford's Law and the detection of election fraud". *Political Analysis*, 19(3):269–272, 2011.

[22] Mebane, W. R. Jr. *Second-digit Tests for Voters' Election Strategies and Election Fraud.* Proceedings of the 2013 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 12–15 2012.

[23] Miller, S. J. *Benford's Law: Theory and Applications.* Princeton University Press, 2015.

[24] Morrow, J. *Benford's Law, Families of Distributions and a Test Basis.* Center for Economic Performance Discussion Paper No. 1291, 2014.

[25] Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. *Amer. J. Math.*, 4.1-4:39–40, 1881.

[26] Nigrini, M. J. A taxpayer compliance application of Benford's Law. *The journal of the American Taxation Association*, 18.1:72–91, 1996.

[27] Nigrini, M. J. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection.* John Wiley & Sons, Hoboken, NJ, 2012.

[28] Nigrini, M. J., Mittermaier, L. J. The use of Benford's law as an aid in analytical procedures. *Auditing: A Journal of Practice & Theory*, 16(2):52–67, 1997.

[29] Nigrini, M. J., Mueller, N. J. Lessons from an $8 million fraud: What the criminal was thinking and what can be done to prevent or uncover similar crimes. *Journal of Accountancy*, 218(2):32–37, 2014.

[30] Pericchi, L., Torres, D. La Ley de Newcomb-Benford y sus aplicaciones al Referendum Revocatorio en Venezuela. Reporte Técnico-definitivo 2a. versión: Octubre 01, 2004.

[31] Pericchi, L., Torres, D. Quick Anomaly Detection by the Newcomb—Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela. *Statistical science*, pages 502–516, 2011.

[32] Pietronero, L., Tosatti, E., Tosatti, V., Vespignani, A. Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf. *Physica A*, 293:297–304, 2001.

[33] Pimbley, J. M. Benford's Law and the Risk of Financial Fraud. *Risk Professional*, May:1–7, 2014.

[34] Pimbley, J. M. Benford's Law as a Logarithmic Transformation. *Maxwell Consulting Archives*, 2014.

[35] Pollach, G., Brunkhorst, F., Mipando, M., Namboya, F., Mndolo, S., Luiz, T. The "first digit law" - A hypothesis on its possible impact on medicine and development aid. *Medical Hypotheses*, 97:102–106, 2016.

[36] Ramini, R. A. The first digit problem. *Amer. Math. Monthly*, 83.7:521–538, 1976.

[37] Rauch, B., Göttsche, M., Brähler, G., Engel, S. Fact and Fiction in EU-Governmental Economic Data. *German Economic Review*, 12(3):243–255, 2011.

[38] Revesz, P. Strong theorems on coin-tossing. *Proc. Internat. Congr. of Mathematicians, Helsinki*, pages 749–754, 1978.

[39] Rücker, G., Freiburg, Uniklinik. Benford's Law in Medical Data Sets. In *49 th Biometric Conference of the German Region of the International Biometric Society at Wuppertal*, 2003.

[40] Sambridge, M., Tkalčić, H., Jackson, A. On the applicability of Benford's Law in the Geosciences. *AGU Fall Meeting Abstracts*, 12 2009.

[41] Shikano, S., Mack, V. When Does the Second-Digit Benford's Law-Test Signal an Election Fraud? Facts or Misleading Test Results. *Jahrbücher für Nationalökonomie und Statistik*, pages 719–732, 2011.

[42] Smith, S. W. *Explaining Benford's Law*. California Technical Pub, The Scientist and Engineer's Guide to Digital Signal Processing, Chapter 34, 2007.

[43] Sottili, G., Palladino, D. M., Giaccio, B., Messina, P. Benford's Law in Time Series Analysis of Seismic Clusters. *Mathematical Geosciences*, 44(5):619–634, 2012.

[44] Tang, Y. Intrinsically disordered proteins(IDPs) obey the Benford's law. *Journal of Baoji University of Arts and Science*, 34(1):1–5, 2015.

[45] Wasserstein, R. L., Lazar, N. A. The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2):129–133, 2016.