[Draft #10] About Distributions of Significant Leading Digits

Vladimir S. Berman
vb7654321@gmail.com

March 4, 2023

Abstract

This paper examines the various forms of distribution densities for leading digits. The formulation of the problem of the distribution of leading digits imposes certain limitations on the Distribution of Significant Leading Digits (DSLD) and should have some specific properties. This provides the possibility of finding an exact solution or alternatively constructing an approximation. This paper presents solutions for the various underlying distributions, with explicit analytical solutions for the first significant digit densities, but for others we find highly accurate approximations. This paper rigorously deduces a new type of DSLD instead of the frequently used empirical NBL.

Contents

1 Notations **3**

2 Introduction **3**

3 Formulation of the Problem **6**

4 Integral representation of DSLD **8**

5 Discrete Sets **10**

6 Continuous Distributions **14**
- 6.1 Examples of Exact Forms of DSLD 14
  - 6.1.1 DSLD of the Ratio of Two Numbers from the UD 16
  - 6.1.2 DSLD of the Production of Two Numbers from a UD . . 17
  - 6.1.3 DSLD of the Production of Three Numbers from a UD . . 18
  - 6.1.4 DSLD of the Production of n Numbers from a UD 19

7 Analysis of DSLD of Other Continuous Distributions **20**
- 7.1 DSLD for LogNormal Distribution 20
- 7.2 DSLD for a First Family of Distributions 21
- 7.3 Examples 23
- 7.4 A Second Family of Distributions 28
  - 7.4.1 DSLD for the Ratio of two Positive Numbers 30
- 7.5 DSLD for Normal Distribution 35
- 7.6 Truncated Normal Distribution 42

8 Approximation **43**
- 8.1 Propertys of $\lfloor x \rfloor$ 43
- 8.2 Discrete Distributions 47
- 8.3 The Fast Fourier Transform. 48
  - 8.3.1 Steler's Law 48
  - 8.3.2 Stigler's FSD Concept 49
- 8.4 Van der Corput's method 50
- 8.5 Weyl Sum 50

9 Discussion **50**

10 Conclusions **52**

A Solution of the Functional Equations **54**
- A.1 Partial solutions 54

B From Sum to Integral in (2) **57**
B.1 Infinite limits 58
B.2 Fourier Expansion 61

C Ratio of Two Positive Numbers with a UD **62**
C.1 Uniform Distributions 63

D The Product of Two Random Positive Numbers from a UD **65**
D.1 Leading significant digits 66

E Product of n Numbers from a UD **67**

F Derivation of Formula (105) **69**

G Useful Relations **69**
G.1 Code 72

List of Figures

1 $a_j = j^\alpha$ 13
2 $a_j = \alpha^j$ 14
3 Exponential Distribution 26
4 Gamma Distributio 27
5 The Ratio of two Positive Numbers 31
6 The ratio of the oscillatory to non-oscilattory parts 32
7 Normal Distribution, $k = 1$ and $\mu = 0$ 38
8 Normal Distribution, $\sigma = 1$ and $\mu = 0$ 39
9 Normal Distribution, $\sigma = 0.1$ and $\mu = 1$ 40
10 Normal Distribution, $\sigma = 1$ and $\mu = -1$ 41
11 $g(x, 1, 10)$ vs x 68

List of Tables

1 Table Caption 1 4
2 Table Caption 2 17
3 Table Caption 3 18
4 Table Caption 4 19
5 Table Caption 5 20
6 Table Caption 6 27
7 Table Caption 7 31
8 Table Caption 8 37
9 Table Caption 9 37
10 Table Caption 10 37
11 Table Caption 11 37
12 Table Caption 12 65

13 Table Caption 13 66

1 Notations

The following notation and definitions will be used in this article.

N	number of elements in set (infinit or finit)
$\rho(k) = \frac{n_k}{N}$	proportion of k in set
DSLD	Distribution of Significant Leading Digits
UD	Uniform Distribution.
$\log x = \dfrac{\ln x}{\ln 10}$	the logarithm of x to base 10.
$\lfloor x \rfloor$	floor, greatest integer less than or equal to a number x.
$x \mod 1$	remainder $\{x\} = x - \lfloor x \rfloor$.
$F = \int\limits_{t=0}^{\infty} f(t)dt$	probability of positive numbers.
$\imath = \sqrt{-1}$	the imaginary unit.
$\rho(k)$	probability mass function (PMF)
CDF	Cumulative Distribution Function
PDF	Probability Density Function
NBL	The Newcomb–Benford Law [15]
$\mathbb{N}$	The set of natural numbers: $\{1, 2, 3, ...\}$
$\mathbb{Z}^*$	The set of non-negative integers: $\{0, 1, 2, 3, ...\}$
$\mathbb{Z}$	The set of integers: $\{..., -3, -2, -1, 0, 1, 2, 3, ...\}$

2 Introduction

For the past 135 years, ever since the problem of the distribution of leading digits began to attract attention [1] and [2], more than $1,500$ articles and books have been published on this subject [10],[16],[18],[19]. Particular attention has been paid to the NBL which describes the DSLD, chosen from empirical distributions of leading digits

$$\rho_{NBL}(k) = \log(k+1) - \log(k), 1 \leq k \leq 9. \tag{1}$$

k	1	2	3	4	5	6	7	8	9
$\rho(k)$	.3010	.1760	.1249	.09690	.07918	.06691	.05799	.05115	.04576

Table 1: Table Caption 1

Many authors believe that the NBL has a universal character and that the majority of cases of number sequences lead to a realization of this law. And in most cases, the distribution of the leading digits based on experimental approaches have been followed by numerical approximations. Often, in the literature, the appearance of the NBL is described as mysterious and magical. Such statements have naturally attracted the attention of various researchers [6],[7],[8],[12],[13],[16],[17],[18] and [19].

An aspiration to use a universal distribution is, in our view, due to the charm of the central limit theorem, where it is proved that for any underlying distribution, we get the normal distribution the in limit. The main attraction of the NBL is due to its simplicity.

However, for the sake of simplicity there is the sacrifice of the important properties of the underlying distributions. So many underlying distributions have parameters whose influence is not reflected in the NBL. Another important feature is the periodic dependence of the DSLD on the parameters underlying the distribution, e.g., for the exponential $(0 < \lambda)$ and normal $(\mu = 0, 0 < \sigma)$ distributions we have, from numerical simulations, a periodic dependence of the DSLD on $\log(\lambda)$ and $\log(\sigma)$, respectively.

Most publications have considered of empirical data sets or numerical interpretation. It is surprising that only a few publications have addressed the underlying probability distributions. In [3] and [7] irregular attempts to use certain symmetry properties have been made to explain the logarithmic nature of the NBL. Symmetry properties are powerful tools that will be used in this paper to find the exact solutions and approximate analysis of the distributions. In our opinion, symmetry plays a very important role in the behavior of the DSLD and is a powerful instrument for analyzing the DSLD. In many cases, the NBL can be regarded as an ad hoc approach to empirical data.

We will build a few basic types of DSLDs. We will find analytical forms of the distribution of leading digits. The main results of this paper concern the presentation of three main constructions of a DSLD.

Let $F(z)$ be a cumulative distribution function (CDF). We will generate a random sample drawn from the distribution given by a continuous random variable Z. We have the distribution function for the sequence of numbers from which we select the sample, and then the density of the leading digits k is defined as

$$10^m k \leq z < 10^m(k+1), \tag{2}$$

$$\rho(k) = \sum_{m=m_{\min}}^{m_{\max}} \left(F(10^m(k+1)) - F(10^m k)\right), \tag{3}$$

$$F(z) = \int\limits_{t=a}^{z} f(t)dt, 10^L \leq k \leq 10^L - 1,$$

$$L \in \mathbb{Z}, L = \lfloor \log(k) \rfloor .$$

Where $f(z)$ is the probability density function (PDF) of the underlying probability distribution, k is the significant $(0 < k)$ leading digit of an element of the sample, and $L = \lfloor \log(k) \rfloor$ defines the order of magnitude of k.

It should be noted that the probability density of the leading digit is not an arbitrary function. It is limited by a number of strong restrictions and symmetries. The explicit form of $\rho(k, L)$ depends on the underlying distribution.

This paper is organized as follows.

The 3 Section discusses the mathematical formulation of the problems. After we solve the functional equations, we have three main types of forms of DSLD.

In Section 4 we give integral relationships for different cases of DSLD.

In Section 5 we consider a few examples of applications of our approach to discrete sets of data.

In the following sections, we present examples capable of exact solutions. In Section 6.1, there are some examples of explicit formulas for $\rho(k)$. More specifically, we present various solutions not only in the interval $1 \ldots 9$ but in the interval $10 \ldots 99$ and so on. We will also discuss how to build analytical solutions for a number of popular underlying distributions. We then build the solutions for this distributions. It is shown that in many cases even a small number of members of sets gives a good approximation of the exact solution. We then proceed, in Sections 7.2,7.4.1 and 7.5 to review the distribution with infinite area of a random sample from an underlying distribution.

Section 9 presents in more detail the assumptions made implicitly. We also present some harmonic and numerical approximations of DSLD which in many cases give exact or 'almost exact' solutions of the problem.

In conclusion, we present examples that show the evolution of different forms of DSLD, and the transition from one form to another when changing the parameters of the underlying distribution. We discuss the characteristics of the DSLD found and forms of correspondence with the NBL.

In the appendices we introduce some technical results useful for understanding the subject of the paper, including integral formulas for the calculation of a DSLD. There we consider in detail many technical aspects of finding solutions and also, for the purpose of self-sufficiency, we have the necessary reference information.

We present a few examples of distributions and show how the properties of the underlying distributions determine the form of the distribution of the first digits. We will consider different underlying distribution, such as, ratio

2 positive numbers from UD, the production of n nonnegative numbers, the distribution of leading digits for an exponential underlying distribution, and ratio 2 positive numbers. In some cases we can derive explicit formulas for the DSLD and trace the evolution of the distribution as a function of the parameters. In other cases, on the basis of a knowledge of the form of the function $\rho(k)$, we will derive approximate forms. Below we give a brief description of the different distributions.

3 Formulation of the Problem

Before formulating the common task, let us consider a simple example. For example, consider the set of two numbers $[0.1, \pi]$, which is equivalent to $[1000\ldots, 31415\ldots]$. Then for $\lfloor \log(k) \rfloor = 0$, first digits are $[1, 3]$, and for $\lfloor \log(k) \rfloor = 1$ the first two digits are $[10, 31]$, and so on.

We will consider a function $\rho(k)$ that has the following symmetry properties. For example, for $k = 1$, we have

$$\begin{aligned}
\rho(1) &= \rho(10) + \rho(11) + \ldots + \rho(19), \\
\rho(1) &= \rho(100) + \rho(101) + \ldots + \rho(199), \\
\rho(1) &= \rho(1000) + \rho(1001) + \ldots + \rho(1999), \\
\rho(10) &= \rho(100) + \rho(101) + \ldots + \rho(109), \\
&\ldots
\end{aligned} \tag{4}$$

In the general case, we have

$$\rho(k) = \sum_{i=0}^{10^L-1} \rho(10^L k + i), L \in \mathbb{Z}_0. \tag{5}$$

$$\begin{aligned}
\rho(k) &= \sum_{i=0}^{10^\delta-1} \rho(10^\delta k + i), \\
&L \in \mathbb{Z}_0, \delta \in \mathbb{Z}_0.
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\sum_{k=10^L}^{10^{L+1}-1} \rho(k, L) = F, L \in \mathbb{Z}_0. \\
&L = \lfloor \log(k) \rfloor, 0 \leq \rho(k, L) \leq 1.
\end{aligned} \tag{7}$$

Here, $0 \leq F$ is the probability of positive numbers.

These requirements put restrictions on the possible forms of the function $\rho(k)$. This raises the interesting question, 'What kind of form should the distribution density of significant leading digits have?'

The relations (6–7) can be interpreted as a system of functional equations. In some cases, these equations have explicit solutions.

For the considered sum, (2) can be expressed through one of the types of construction from the list below.

Some solutions of the functional equations (6–7) are in Appendix A.

- In case $\rho(k, L)$ only depends on L, then all the digit for the same L have the same probability, and we have

$$\rho(k) = \frac{1}{9} 10^{-\lfloor \log(k) \rfloor}. \tag{8}$$

- In case $\rho(k, L)$ depends only on k, the solution is

$$\begin{aligned} \rho(k) &= F\left(\log(k+1) - \log(k)\right) \\ &\quad + Q_1(\log(k+1)) - Q_1(\log(k)), \\ Q_1(s) &= Q_1(s+1), \forall s. \end{aligned} \tag{9}$$

- In the more general case, when $\rho(k, \lfloor \log(k) \rfloor)$, we have

$$\rho(k) = \Omega(\log(k+1) - \lfloor \log(k) \rfloor) - \Omega(\log(k) - \lfloor \log(k) \rfloor). \tag{10}$$

Here, $\Omega(s)$ is any admissible function.

In the case when $j_{\max} = +\infty$ and $j_{\min} = -\infty$ we can see that the function $G(k) = const + \sum_{j=-\infty}^{\infty} F(10^j k)$ is invariant under transformation with integer a

$$j \to j' - a, \log(k) \to \log(k' 10^a) = \log(k') + a.$$

In this case we have a solution in the form of the sum of a 1-period function and a linear function.

$$const + \sum_{j=-\infty}^{\infty} F(10^{j+\log(k)}) = const + A\log(k) + Q_1(\log(k)),$$
$$Q_1(s+1) = Q_1(s), \forall s.$$

Sometimes it is convenient to add, without loss of generality, to the argument of Q_1, the integer $-\lfloor \log(k) \rfloor$

$$const + \sum_{j=-\infty}^{\infty} F(10^{j+\log(k)}) = A\log(k) + Q_1(\log(k) - \lfloor \log(k) \rfloor).$$

and (9) takes the form

$$\begin{aligned} \rho(k) = A\left(\log(k+1) - \log(k)\right) + Q_1(\log(k+1) \\ - \lfloor \log(k) \rfloor) - Q_1(\log(k) - \lfloor \log(k) \rfloor). \end{aligned} \tag{11}$$

$$\log(k) - \lfloor \log(k) \rfloor \equiv \log(k) \mod 1 = \{k\}.$$

From (10) we can present (8) as

$$\Omega(s) = \frac{1}{9}10^s,$$

and (9) as

$$\Omega(s) = F \log(s) + Q_1(s). \tag{12}$$

We can enter a condition in which the implemented solution (9) for all permissible L and k

$$\Omega(\log(k+1)) - \Omega(\log(k+1) + L) \tag{13}$$
$$-\Omega(\log(k)) + \Omega(\log(k) + L) = 0,$$
$$k \in \mathbb{N}, L \in \mathbb{Z}^*.$$

4 Integral representation of DSLD

Because operate with discrete relations, which are less convenient than continuous ones, it would be useful to replace the summation in (2) with an integration. We use the well-known Euler–Maclaurin formula (EMf)[4],[5] and [9] and make an exact replacement of the sum by the integral.Euler–Maclaurin's formula (EMf) gives a representation of a sum in the form of a definite integral. It is important to stress that the EMf is an exact expression of the sum in integral form (for at least a differentiable function): it is not an approximation. From the EMf we have an expression the sum in the form of an integral which is (see

Appendix F)

$$\sum_{j=a}^{b} g(j) = \int\limits_{x=a}^{b} \left(g(x) + \left(x - \lfloor x \rfloor - \frac{1}{2} \right) \frac{dg(x)}{dx} \right) dx \tag{14}$$
$$+ \frac{g(b)}{2} + \frac{g(a)}{2}, \tag{15}$$

a and b are integers, $a < b$.

We will apply Euler–Maclaurin summation in this form, which is convenient for our purposes.

In our case we have

$$g(j) = F(10^{j+\log(k+1)}) - F(10^{j+\log(k)})$$
$$F(z) = \int\limits_{-\infty}^{z} f(t)dt.$$

Depending on the values of a and b, we have different representations (14). When $a = -\infty (z = 0)$ and $b = \infty (z = \infty)$ we have

$$\rho(k) = \int\limits_{t=0}^{\infty} f(t) P(k,t) dt, \tag{16}$$

$$\Omega(s) = -\int\limits_{t=0}^{\infty} f(s) \left\lfloor \log(t) - s \right\rfloor dt. \tag{17}$$

where

$$P(k,t) = \left\lfloor \log\left(\frac{t}{k}\right) \right\rfloor - \left\lfloor \log\left(\frac{t}{k+1}\right) \right\rfloor. \tag{18}$$

For a positive integer k and a positive t we have

$$P(k,t) = \begin{cases} 1, & 10^m k \le t \le 10^m (k+1), \\ 0, & t \notin [10^m k, 10^m (k+1)], \text{integer } m. \end{cases} \tag{19}$$

Formula (16) is equivalent to

$$\rho(k,L) = \sum_{j=-\infty}^{\infty} \int\limits_{10^{j-L}k}^{10^{j-L}(k+1)} f(t) dt, \tag{20}$$
$$L = \left\lfloor \log(k) \right\rfloor.$$

In the case when the function $f(t)$ has no fixed points in the interior of the interval $[t_{\min}, t_{\max}]$ or $-\infty < j_{\min}$ or/and $j_{\max} < \infty$ then we have a solution independent of $\lfloor \log(k) \rfloor$ (9). Conveniently, in the spirit of (9), this case can be expressed as the sum of its logarithmic and periodic members

$$\rho(k) = F\left(\log(k+1) - \log(k)\right) + Q_1(\log(k+1)) - Q_1(\log(k+1)), \tag{21}$$

$$Q_1(\log(k)) = \int\limits_{s=0}^{\infty} f(s) \left(\log\left(\frac{s}{k}\right) - \left\lfloor \log\left(\frac{s}{k}\right) \right\rfloor \right) ds, \tag{22}$$

$$F = \int\limits_{s=0}^{\infty} f(s) ds. \tag{23}$$

It is interesting to note that the first term in (21) has the form for any admissible function $f(s)$, this follows from integration by parts.

$$\int\limits_{x=-\infty}^{\infty} \int\limits_{t=10^x k}^{10^x (k+1)} f(t) dt = \left(\log(k+1) - \log(k)\right) \int\limits_{t=0}^{\infty} f(t) dt. \tag{24}$$

The function $Q_1(s)$ is 1-periodic in its argument. It is natural to consider the Fourier decomposition of this function. Then, after a few operations (see Appendix B.2), we can write $\rho(k)$ as

$$\rho(k) = F\left(\log(k+1) - \log(k)\right) + \sum_{n=1}^{\infty}\Lambda(n)\left(\sin\left(ns_1+\varphi(n)\right) - \sin\left(ns+\varphi(n)\right)\right), \tag{25}$$

$$s_1 = 2\pi\log(k+1), s_1 = 2\pi\log(k), F = \int\limits_{x=0}^{\infty} f(x)dx \tag{26}$$

$$\Lambda(n)\exp(\imath\varphi(n)) = \frac{1}{n\pi}\int\limits_{x=0}^{\infty} x^{-\imath Y} f(x)dx, Y = \frac{2\pi n}{\ln(10)}. \tag{27}$$

In many practical cases, the Fourier coefficients decrease rapidly and then it is possible to confine one's attention to only a few (even the first one) of the first ones.

5 Discrete Sets

Consider a set of N positive numbers. It is convenient to introduce an enumeration of the elements of this set (the order is not important). Then

$$f(z,N) = \frac{1}{N}\sum_{j=1}^{N}\delta\left(z-a_j\right). \tag{28}$$

From (111)

$$P(k,x) = \left\lfloor \log\left(\frac{x}{k}\right)\right\rfloor - \left\lfloor \log\left(\frac{x}{k+1}\right)\right\rfloor. \tag{29}$$

we have

$$\rho(k,N) = \frac{1}{N}\sum_{j=1}^{N}P(k,a_j).$$

And hence we obtain

$$\Omega(s) = -\frac{1}{N}\sum_{j=1}^{N}\left\lfloor\log\left(a_j\right) - s\right\rfloor + const, \tag{30}$$

$$\Omega(1) - \Omega(0) = 1. \tag{31}$$

This is the general solution. But in some cases $\rho(k,N)$ can monotonically go to the NBL or oscillate around the NLB. In such a case, it is more convenient to use the form (121) and (122). Then

$$A_n(N)+\imath B_n(N)=\frac{1}{n\pi}\int\limits_{x=0}^{\infty}x^{-\imath Y}f(x,N)dx, Y=\frac{2\pi n}{\ln(10)},$$

ant taking in consideration (28) we have

$$A_n(N)+\imath B_n(N)=\frac{1}{n\pi N}\sum_{j=1}^{N}(a_j)^{-\imath Y} \tag{32}$$

$$=\Lambda(n,N)\exp(\imath\varphi(n,N)), Y=\frac{2\pi n}{\ln(10)}, \tag{33}$$

and

$$\rho(k,N)=\log(k+1)-\log(k)+ \tag{34}$$
$$\sum_{n=1}^{\infty}\begin{pmatrix} A_n\left(\sin(ns_1)-\sin(ns)\right)\\ +B_n(\cos(ns_1)-\cos(ns))\end{pmatrix} \text{ or}$$

$$\rho(k,N)=\log(k+1)-\log(k)+\sum_{n=1}^{\infty}\Lambda(n,N)\left(\sin(ns_1+\varphi(n,N))-\sin(ns+\varphi(n,N))\right), \tag{35}$$

$$s_1=2\pi\log(k+1), s=2\pi\log(k).$$

Let us introduce a criterion for assessing the contribution of the oscillatory term

$$V(k,n,N)=\left|\frac{\sum\limits_{j=1}^{n}\Lambda(j,N)\left(\sin(js_1+\varphi(j,N))-\sin(js+\varphi(j,N))\right)}{\log\left(1+\frac{1}{k}\right)F}-1\right|, F=1. \tag{36}$$

if $V\ll 1$ then this is the case of the NBL. In many cases we need only a few terms in the sum ($n=1,2,3,\dots$).

Let us consider two examples.

- $a_j=\alpha^j$

$$\frac{1}{n\pi N}\sum_{j=1}^{N}(a_j)^{-\imath Y}=\frac{1}{n\pi N}\frac{1-\exp(-N\beta\imath)}{\exp(\beta\imath)-1}, \beta=2\pi n\log(\alpha). \tag{37}$$

If $\alpha = 10^r$ where $r = \pm 1, \pm 2, \ldots$, then the RHS is

$$\beta = 2\pi n r, \Lambda(n, N) = \frac{1}{n\pi}, B_n(N) = 0,$$
$$\rho(k, N) = \begin{cases} 1, & k = 1, \\ 0, & k \neq 1, k < 10 \end{cases} .$$

If $\alpha \neq 10^r$, then

$$\Lambda(n, N) = \frac{1}{\pi n N}\sqrt{\frac{1 - \cos(\beta N)}{1 - \cos(\beta)}},$$
$$\tan(\varphi(n, N)) = -\frac{\sin(\beta) + \sin(N\beta)}{\cos(\beta) + \cos(N\beta)}$$

How can we see that $\rho(k, N)$ as $N \to \infty$ approaches the NBL, with damped oscillations.

- $a_j = j^\alpha$ where $\alpha \neq 0$ is real

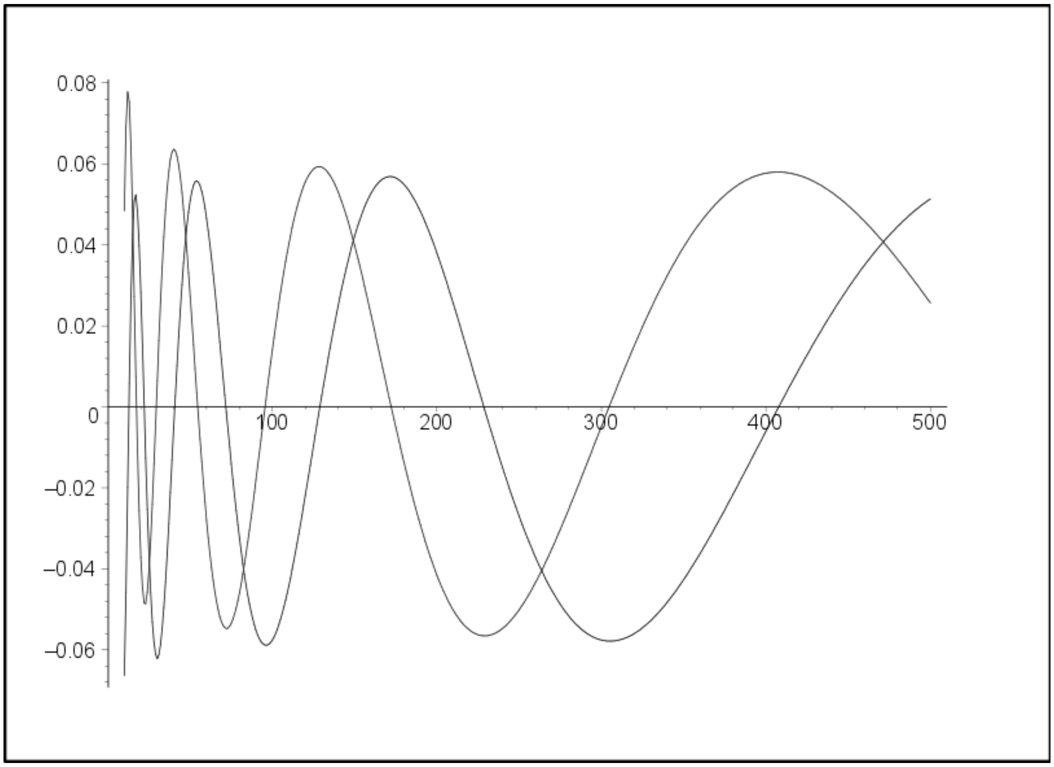
$$\frac{1}{n\pi N}\sum_{j=1}^{N}(a_j)^{-\imath Y} = \frac{1}{n\pi N}\sum_{j=1}^{N}\exp(-\imath\beta\log(j)), \beta = 2\pi n\alpha.$$

When $N \to \infty$ we can convert the summation to an integration

$$\frac{1}{n\pi N}\sum_{j=1}^{N}\exp(-\imath\beta\log(j)) \approx \frac{1}{n\pi N}\int\limits_{x=1}^{N}\exp(-\imath\beta\log(x))dx \tag{38}$$
$$\approx \frac{1}{n\pi}\exp(-\imath\beta\log(N))\int\limits_{1/N}^{1}\exp(-\imath\beta\log(y))dy$$

from here we have that $\rho(k, N \to \infty)$ is a sum of oscillating functions from $\log(N)$ with periods $T_n = \frac{1}{n|\alpha|}$

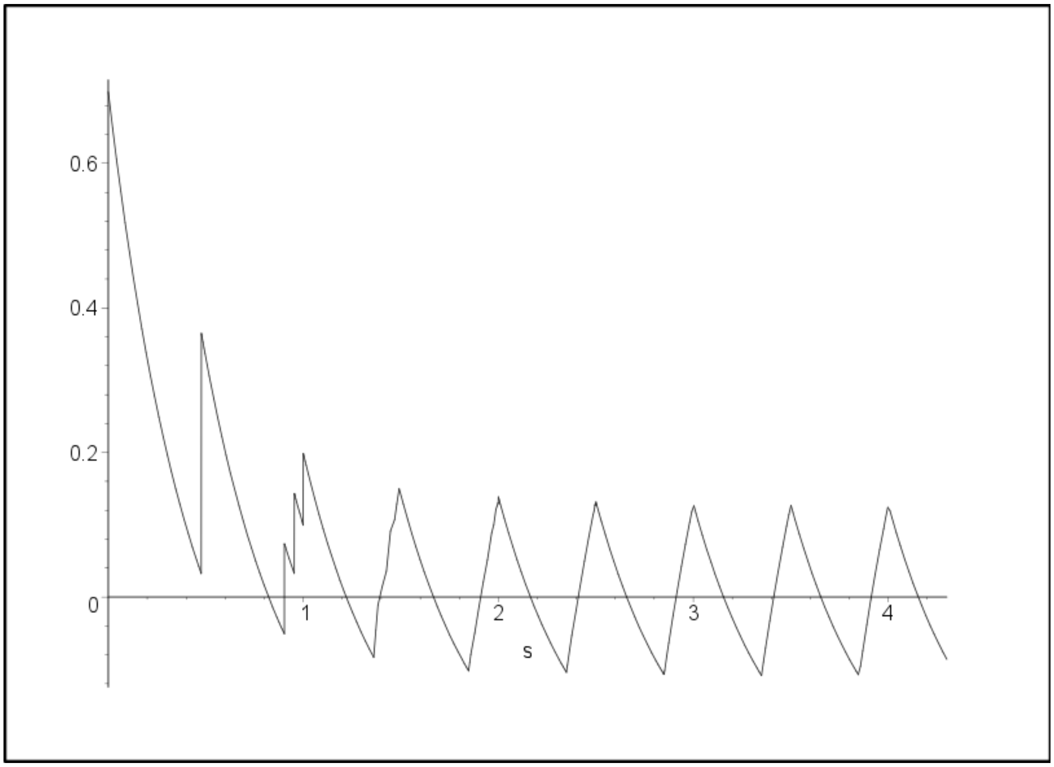
Numerical calculation give us the dependencies of $A_n(N)$ and $A_n(N)$ on N



$a_j = j^{-2}, A_n(N)$ - green line, $B_n(N)$ - blue line

Figure 1: $a_j = j^{\alpha}$

and



$a_j = j^{-2}, T_1 = \frac{1}{2}, \rho(k=1,N) - \log(2)$ vs $s = \log(N)$

Figure 2: $a_j = \alpha^j$

and $V(k = 1, n = 1, N = 10^4, = .2793$, so the contribution of the oscillatory term is substantial.

6 Continuous Distributions

6.1 Examples of Exact Forms of DSLD

This section and the next explore simple examples of distributions and find explicit forms of DSLD, giving a direction for

what's to come. They have a future in common: all these problems involve explicit analytical distributions of the leading digits.
Their treatments all use the idea of the solution of the basic equations (2,7,5,6).

- DSLD **(8)** for **Power Distribution** ($b = 1$). One example can be

$$PDF = \begin{cases} nz^{n-1}, & 0 \le z \le 1, 0 \le n, \\ 0, & \text{otherwise} \end{cases}$$
$$F(z) = \begin{cases} z^n, & 0 \le z \le 1, 0 \le n, \\ 0, & \text{otherwise} \end{cases}.$$

Then we have, from (2),

$$\rho(k) = \sum_{m=-\infty}^{m_{\max}} \left(F(10^m(k+1)) - F(10^m k)\right).$$

and from the condition

$$10^{m_{\max}} k < 1,$$
$$m_{\max} = -\lfloor \log(k) \rfloor - 1.$$

$$\rho(k,n) = \sum_{m=-\infty}^{-\lfloor \log(k) \rfloor - 1} 10^{mn} \left((k+1)^n - k^n\right).$$

$$\begin{aligned} s &= \log(k) - \lfloor \log(k) \rfloor, \\ s_1 &= \log(k+1) - \lfloor \log(k) \rfloor, \\ \rho(k,n) &= \frac{10^{ns_1} - 10^{ns}}{10^n - 1}, \\ \Omega(s) &= \frac{10^{ns}}{10^n - 1}, \Omega(1) - \Omega(0) = 1. \end{aligned} \tag{39}$$

If $n = 1$ then

$$\rho(k,1) = \frac{10^{-\lfloor \log(k) \rfloor}}{9}.$$

When $n \to 0$ we have the NBL:

$$\rho(k, n \to 0) = \log(k+1) - \log(k) + O(n).$$

This distribution describes not just the first digits but it is good for a collection of first digits. For example, for $k = 123$,

$$\begin{aligned} \rho(123,n) &= \frac{1.24^n - 1.23^n}{10^n - 1}, \\ \rho(123, 0.5) &= .00185, \\ \rho(123, 1.0) &= .001111, \\ \rho(123, 2.0) &= .00021. \end{aligned}$$

- Here we can consider similar **Pareto distribution** ($b = 1$) with

$$CDF = 1 - z^{-n}, 0 < n, 1 \leq z \leq \infty.$$

with

$$\begin{aligned} s &= \log(k) - \lfloor \log(k) \rfloor, \\ s_1 &= \log(k+1) - \lfloor \log(k) \rfloor, \\ \rho(k,n) &= \frac{10^n \left(-10^{-ns_1} + 10^{-ns}\right)}{10^n - 1}, \\ \Omega(s) &= -\frac{10^{n-ns}}{10^n - 1}, \Omega(1) - \Omega(0) = 1.. \end{aligned} \tag{40}$$

and

$$\rho(k,1) = \frac{10^{1+\lfloor \log(k) \rfloor}}{9} \left(\frac{1}{k} - \frac{1}{k+1} \right),$$

with convergence to the NBL as $n \to 0$

$$\rho(k, n \to 0) = \log(k+1) - \log(k) + O(n).$$

From (39 and 40) we have for $\Omega(s)$ from (13) a transition to NBL with $n = 0$.

6.1.1 DSLD of the Ratio of Two Numbers from the UD

From Appendices A and C we have

$$F(z) = \begin{cases} \frac{z}{2} & 0 \leq z \leq 1, \\ 1 - \frac{1}{2z} & 1 < z. \end{cases}$$

$$\rho(k,0) = \sum_{m=-\infty}^{-1} \left(F(10^m(k+1)) - F(10^m k) \right), 1 \leq k \leq 9.$$

$$\rho(k,0) = \frac{1}{18} - \frac{5}{9} \left(\frac{1}{k+1} - \frac{1}{k} \right), 1 \leq k \leq 9. \tag{41}$$

$$\begin{aligned} \Omega(s) &= A10^{-s} + B10^{s}, \\ A &= -\frac{5}{9}, B = \frac{1}{18}, \end{aligned}$$

$\Omega(1) - \Omega(0) = 1.$

$$\begin{aligned} s_1 &= \log(k+1) - \lfloor \log(k) \rfloor, \\ s &= \log(k) - \lfloor \log(k) \rfloor. \end{aligned}$$

k	1	2	3	4	5	6	7	8	9
$\rho(k,0)$	0.3333	0.1481	0.1019	0.08333	0.07407	0.06878	0.065488	0.06327	0.06173

Table 2: Table Caption 2

$$\rho(k,L) = \Omega(s_1) - \Omega(s), k \text{ any positive integer}$$

$$\rho(k,L) = \frac{10^{-L}}{18} - \frac{5}{9}\left(\frac{1}{k+1} - \frac{1}{k}\right)10^L, L = \lfloor \log(k) \rfloor, 1 \leq k \leq 9. \tag{42}$$

We can find the distribution of the second digits:

$$\varrho_2(m) = \sum_{i=1}^{9} \rho(10i+m,1),$$

$$\varrho_2(m) = \sum_{i=1}^{9}\left(\frac{1}{180} + \frac{50}{9(10i+m)} - \frac{50}{9(10i+m+1)}\right), m = 0,1,\dots,9. \tag{43}$$

$\varrho_2(0) = .1294, \varrho_2(1) = .1191, \varrho_2(2) = .1109, \varrho_2(5) = .09423, \varrho_2(9) = .08169.$
and the distribution of the third digits p:

$$\varrho_3(p) = \sum_{i=1}^{9}\sum_{j=0}^{9} \rho(100i+10j+p,2),$$

$$\varrho_3(p) = \sum_{i=1}^{9}\sum_{j=0}^{9}\left(\frac{1}{1800} + \frac{500}{9(100i+10j+p)} - \frac{500}{9(100i+10j+p+1)}\right), \tag{44}$$

$$p = 0,1,\dots,9. \tag{45}$$

$\varrho_3(0) = .10254, \varrho_3(9) = .09759.$

We have what is almost the uniform distribution with a density close to 0.1.

6.1.2 DSLD of the Production of Two Numbers from a UD

From Appendix D

$$f_Z(z,2) = -\ln(z), 0 \leq z \leq 1.$$

$$F_Z(z,2) = z\left(1 - \ln(z)\right), 0 \leq z \leq 1.$$

$$\rho(k, L=0, 2) = \sum_{m=-\infty}^{m=-1}\left(F_Z(10^m(k+1),2) - F_Z(10^m k),2\right), 1 \leq k \leq 9.$$

$$\Omega(s) = 10^s \left(A + Bs\right),$$
$$A = \frac{1}{9} + \frac{10}{81}\ln(10), B = -\frac{\ln(10)}{9}.$$

$$s_1 = \log(k+1) - \lfloor \log(k) \rfloor,$$
$$s = \log(k) - \lfloor \log(k) \rfloor.$$

$$\rho(k, L, 2) = \Omega(s_1) - \Omega(s),$$
$$\Omega(1) - \Omega(0) = 1, k \text{ any positive integer.}$$

$$\rho(k, L=0, 2) = \frac{1}{9}\left(k\ln(k) - (k+1)\ln(k+1) + 1 + \frac{10}{9}\ln(10)\right), 1 \le k \le 9. \tag{46}$$

k	1	2	3	4	5	6	7	8	9
$\rho(k,0,2)$	0.24135	0.18319	0.14551	0.11736	0.09500	0.07640	0.06050	0.04660	0.03410

Table 3: Table Caption 3

$$\rho(k, L, 2) = \frac{10^{(-L)}}{9}\left(k\ln(k) - (k+1)\ln(k+1) + 1 + \frac{(9L+10)}{9}\ln(10)\right), \tag{47}$$
$$10^L \le k \le 10^{L+1} - 1, L = 0, 1, \ldots$$

6.1.3 DSLD of the Production of Three Numbers from a UD

From Appendix E we have

$$f_Z(z, 3) = \frac{\ln(z)^2}{2}, 0 \le z \le 1.$$

$$F_Z(z, 3) = \frac{z}{2}\left(\ln(z)^2 - 2\ln(z) + 2\right), 0 \le z \le 1.$$

$$\begin{aligned}\rho(k, 0) = & \frac{1}{18}\left((k+1)\ln(k+1)^2 - k\ln(k)^2\right) \\ & - \left(\frac{1}{9} + \frac{10}{81}\ln(10)\right)\left((k+1)\ln(k+1) - k\ln(k)\right) \\ & + \frac{55}{729}\left(\ln(10)\right)^2 + \frac{10}{81}\ln(10) + \frac{1}{9}\end{aligned}$$

k	1	2	3	4	5	6	7	8	9
$\rho(k,0)$	0.30066	0.18817	0.13194	0.09864	0.07699	0.06290	0.05260	0.04630	0.04179

Table 4: Table Caption 4

$$\begin{aligned}
\Omega(s) &= 10^s \left(A + Bs + Cs^2\right),\\
C &= \frac{\ln(10)^2}{18}, B = -\frac{\ln(10)}{81}(9 + 10\ln(10)),\\
A &= \frac{1}{9} + \frac{10}{81}\ln(10) + \frac{55}{729}\ln(10)^2
\end{aligned}$$

$$\Omega(1) - \Omega(0) = 1.$$

$$\begin{aligned}
s_1 &= \log(k+1) - \lfloor \log(k) \rfloor,\\
s &= \log(k) - \lfloor \log(k) \rfloor.
\end{aligned}$$

$$\rho(k) = \Omega(s_1) - \Omega(s), k \text{ any positive integer}$$

6.1.4 DSLD of the Production of n Numbers from a UD

From Appendix E

$$\begin{aligned}
z &= \prod_{i=1}^{n} x_i,\\
z &\le x_1 \le 1,\\
x_n &= z \left(\prod_{i=1}^{n-1} x_i\right)^{-1},
\end{aligned}$$

$$z \left(\prod_{i=1}^{j-1} x_i\right)^{-1} \le x_j \le 1, j = 2, \ldots, n-1,$$

$$J = \frac{\partial(x_1, \ldots, x_n)}{\partial(x_1, \ldots, z)} = \frac{\partial x_n}{\partial z} = \left(\prod_{i=1}^{n-1} x_i\right)^{-1}.$$

$$f_Z(z,n) = \int_{x_1=z}^{1} \int_{x_2=z/x_1}^{1} \cdots \int_{x_{n-1}=z/(x_1 \ldots x_{n-2})}^{1} J^{-1} dx_{n-1} \ldots dx_2 dx_1, 2 \le n.$$

$$f_Z(z,n) = \frac{(-\ln(z))^{(n-1)}}{(n-1)!}, 0 \le z \le 1.$$

$$F_Z(z,n) = \frac{\Gamma(n, -\ln(z))}{\Gamma(n)}, 0 \le z \le 1.$$

where

$$\Gamma(a,x) = \int_x^{\infty} t^{a-1}\exp(-t)dt,$$

$$\Gamma(a) = \int_0^{\infty} t^{a-1}\exp(-t)dt = (a-1)!.$$

Then as $n \to +\infty$, we have, from Appendix E

$$\Omega\left(s\right) = s + const,$$
$$\rho(k, n \to +\infty) = \log(k+1) - \log(k).$$

This is the distribution of the NBL. The table presents the values of $\rho(k, n \to +\infty)$ for some parameters

n	2	3	4	5	6	7	8	9
$\rho(1,n)$	.24135	.30066	.30764	.30279	.30068	.30074	.30100	.30106
$\rho(9,n)$	.03418	.04167	.04585	.04619	.04585	.04573	.04573	.04573

Table 5: Table Caption 5

$$\log(2) = 0.30103, 1 - \log(9) = 0.045757.$$

7 Analysis of DSLD of Other Continuous Distributions

7.1 DSLD for LogNormal Distribution

$$f(z) = \frac{1}{\sigma z\sqrt{2\pi}}\exp\left(-\frac{\left(\ln(z)-\mu\right)^2}{2\sigma^2}\right),$$

$$F(z) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\left(\ln(z)-\mu\right)\sqrt{2}}{2\sigma}\right), 0 \le z < \infty.$$

Then

$$Q_1\left(s,\sigma\right) = const + \sum_{n=1}^{\infty}\left(A_n\sin\left(ns\right) + B_n\cos\left(ns\right)\right). \tag{48}$$

$$F = 1,$$
$$A_n = \frac{1}{n\pi}\exp\left(-\frac{2\pi^2\sigma^2}{\ln(10)^2}n^2\right), B_n = 0.$$

$$\rho(k,\mu,\sigma\to 0) = \log(k+1) - \log(k) + Q_1(s_1,\sigma) - Q_1(s,\sigma). \quad (49)$$
$$s_1 = 2\pi\left(\log(1+k) - \frac{\mu}{\ln(10)}\right), s = 2\pi\left(\log(k) - \frac{\mu}{\ln(10)}\right).$$

From this we obtain that $\rho(k,\mu,\sigma)$ is a periodic function of μ with period $\ln(10)$.

As $\sigma \to 0$ ($\mu = 0$) we have

$$\rho(k) = \Omega(\log(k+1) - \lfloor\log(k)\rfloor) - \Omega(\log(k) - \lfloor\log(k)\rfloor)$$
$$\rho(k) = \begin{cases} \frac{1}{2}, & \log(k) = \lfloor\log(k)\rfloor, \\ \frac{1}{2}, & \log(k+1) = \lfloor\log(k)\rfloor + 1. \end{cases}$$

Where

$$\Omega(x) = \begin{cases} -\frac{1}{2}, & x = 0, \\ \frac{1}{2}, & x = 1, \\ 0, & \text{other}. \end{cases},$$
$$\Omega(1) - \Omega(0) = 1.$$

As a result of the summation,

$$Q_1(s,\sigma) = A_0 + \arctan\left(\frac{\beta\sin(2\pi s)}{1 - \beta\cos(2\pi s)}\right), \beta = \exp\left(-\frac{2\pi^2\sigma^2}{\ln(10)^2}\right). \quad (50)$$

If we compare the results of our approach and the numerical results of stochastic modeling [15]. We can see that despite the fact that it was used in the numerical modeling of 10^7 samples, the accuracy of the approximation is quite poor. The approximation errors lie in the range of $15\% - 100\%$.

7.2 DSLD for a First Family of Distributions

Here we will look at one type of density function, that underlying distributions that describe a number of important and frequently used distributions [14] .

$$F = \int_{z=0}^{\infty} f(z)dz.$$

The model distribution is a continuous probability distribution with probability density function given by

$$f(z) = \begin{cases} 0, & z < 0 \\ F\beta\left(\frac{z}{a}\right)^{\alpha}\exp\left(-\left(\frac{z}{a}\right)^{\beta}\right)a^{(-1)}\Gamma\left(\frac{\alpha+1}{\beta}\right)^{-1}, & 0 \le z, 0 < a, 0 < \alpha, 0 < \beta. \end{cases}$$

$$ \tag{51} $$

Because from general form we get

$$\rho(k,a,\alpha,\beta) = \sum_{m=-\infty}^{\infty} \left(F(\frac{10^m(k+1)}{a} - F\left(\frac{10^m k}{a}\right)\right). \tag{52}$$

From the leading term on the right hand side of (52) we have

$$\begin{aligned}
&m + \log(k) - \log(a) \\
&= m + \lfloor \log(k) \rfloor - \lfloor \log(a) \rfloor \\
&+ \log(k) - \lfloor \log(k) \rfloor - \log(a) + \lfloor \log(a) \rfloor
\end{aligned}$$

and we have (52) with an infinite range of summation. Now we can introduce a new dummy variable m'

$$m = m' - \lfloor \log(k) \rfloor + \lfloor \log(a) \rfloor.$$

After some calculation we have

$$\begin{aligned}
\rho(k,a,\alpha,\beta) &= \log(k+1) - \log(k) + \\
&\sum_{n=1}^{\infty} A_n(\alpha,\beta)\left(\sin(ns_1) - \sin(ns)\right) + \\
&\sum_{n=1}^{\infty} B_n(\alpha,\beta)\left(\cos(ns_1) - \cos(ns)\right), \\
s_1 &= 2\pi \log\left(\frac{k+1}{a}\right), s = 2\pi \log\left(\frac{k}{a}\right).
\end{aligned}$$

Where

$$A_n(\alpha,\beta) + iB_n(\alpha,\beta) = F\frac{1}{\pi n}\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta} - Y\imath\right)\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta}\right)^{-1}, \tag{53}$$

$$Y = \frac{2\pi n}{\beta \ln(10)}, \tag{54}$$

From this we obtain that $\rho(k,a,\alpha,\beta)$ is a periodic function of $\log(a)$ with period 1.

A useful property of the Γ function is

$$\frac{|\Gamma(x + \imath y)|}{|\Gamma(x)|} \leq 1, \text{ for any real } x \text{ and } y. \tag{55}$$

then

$$\sqrt{A_n(\alpha,\beta)^2 + B_n(\alpha,\beta)^2} \leq \frac{F}{\pi n}. \tag{56}$$

7.3 Examples

- **Exponential Distribution**

$$f(z,\lambda) = \frac{1}{\lambda}\exp\left(-\frac{z}{\lambda}\right), 0 < \lambda, 0 \le z.$$

$$\begin{aligned}
F &= 1, \alpha = 0, \beta = 1,\\
A_n + iB_n &= \frac{1}{\pi n}\Gamma\left(1 - \imath Y\right)\Gamma\left(1\right)^{-1},\\
\Lambda(n)^2 &= 2\left(n\ln(10)\sinh\left(\frac{2\pi^2 n}{\ln(10)}\right)\right)^{-1},\\
s_1 &= 2\pi\log\left(\frac{k+1}{\lambda}\right), s = 2\pi\log\left(\frac{k}{\lambda}\right).
\end{aligned}$$

$\Lambda(1) = .01813, A_1 = .01308, B_1 = -.01256,$
$\Lambda(2) = 1.768\times 10^{-4}, A_2 = 2.484\times 10^{-5}, B_2 = 1.749\times 10^{-4}.$

- **Weibull Distribution**

$$f(z,\nu) = \frac{\beta}{a}\left(\frac{z}{a}\right)^{\beta-1}\exp\left(-\left(\frac{z}{a}\right)^{\beta}\right), 0 \le z.$$

$$\begin{aligned}
F &= 1, \alpha = \beta - 1, 0 < \beta,\\
A_n(\beta) - iB_n(\beta) &= \frac{2i}{\beta\ln(10)}\Gamma\left(1 + \imath\frac{Y}{\beta}\right),\\
\Lambda(n,\beta)^2 &= \frac{2}{\beta\ln(10)n}\sinh\left(\frac{2\pi^2 n}{\beta\ln(10)}\right)^{-1},\\
s_1 &= 2\pi\log\left(\frac{k+1}{a}\right), s = 2\pi\log\left(\frac{k}{a}\right).
\end{aligned}$$

$\Lambda(1,2) = .1093, A_1(2) = .1068, B_1(2) = .02362,$
$\Lambda(2,2) = 9.065\times 10^{-3}, A_2(2) = .006552, B_2(2) = -.006283.$

$$\begin{aligned}
\rho(k,a,n) &= \log(k+1) - \log(k)\\
&\quad + \sum_{j=1}^{n} A_j(\sin(js_1) - \sin(js))\\
&\quad + \sum_{j=1}^{n} B_j(\cos(js_1) - \cos(js)),\\
s_1 &= 2\pi\frac{\log(k+1)}{a}, s = 2\pi\frac{\log(k)}{a}.
\end{aligned}$$

with $1 \ll \beta$ we have

$$A_n(\beta) = \frac{1}{\pi n} + o\left(\frac{1}{\beta}\right),$$
$$B_n(\beta) = \frac{2\gamma}{\beta \ln(10)} + o\left(\frac{1}{\beta^2}\right), \gamma = .57721$$

- **ChiSquare Distribution**

$$f(z,\nu) = \left(\frac{z}{2}\right)^{\frac{\nu}{2}-1} \exp\left(-\frac{z}{2}\right)\left(2\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}, 0 \le z.$$

$$F = 1, a = 2, \alpha = \frac{\nu}{2} - 1, \beta = 1,$$

$$A_n(\nu) - iB_n(\nu) = \frac{1}{\pi n}\Gamma\left(\frac{\nu}{2} + Y\imath\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}, \tag{57}$$
$$Y = \frac{2\pi n}{\ln(10)}, \tag{58}$$
$$s_1 = 2\pi \log\left(\frac{k+1}{2}\right), s = 2\pi \log\left(\frac{k}{2}\right). \tag{59}$$

$$A_1(7) = -.1041, B_1(7) = .01788,$$
$$A_1(7) = .0005160, B_1(7) = -.004204.$$

$$\rho(k,a,n) = \log(k+1) - \log(k)$$
$$+ \sum_{j=1}^{n} A_j(\sin(js_1) - \sin(js))$$
$$+ \sum_{j=1}^{n} B_j(\cos(js_1) - \cos(js)),$$
$$s_1 = 2\pi\frac{\log(k+1)}{2}, s = 2\pi\frac{\log(k)}{2}.$$

With $1 \ll \nu$

$$A_n(\nu) = \frac{1}{\pi n}\cos\left(2\pi n \log\left(\frac{\nu}{2}\right)\right),$$
$$B_n(\nu) = -\frac{1}{\pi n}\sin\left(2\pi n \log\left(\frac{\nu}{2}\right)\right).$$
$$Q_1(k, 1 \ll \nu) = \sum_{n=1}^{\infty}\frac{1}{\pi n}\sin\left(2\pi n \log\left(\frac{k}{\nu}\right)\right)$$

$$\sum_{n=1}^{\infty} \frac{1}{\pi n} \sin(2\pi n X) = \frac{1}{\pi} \arctan\left(\frac{\sin(2\pi X)}{1-\cos(2\pi X)}\right)$$
$$= \frac{1}{2} - (X - \lfloor X \rfloor).$$

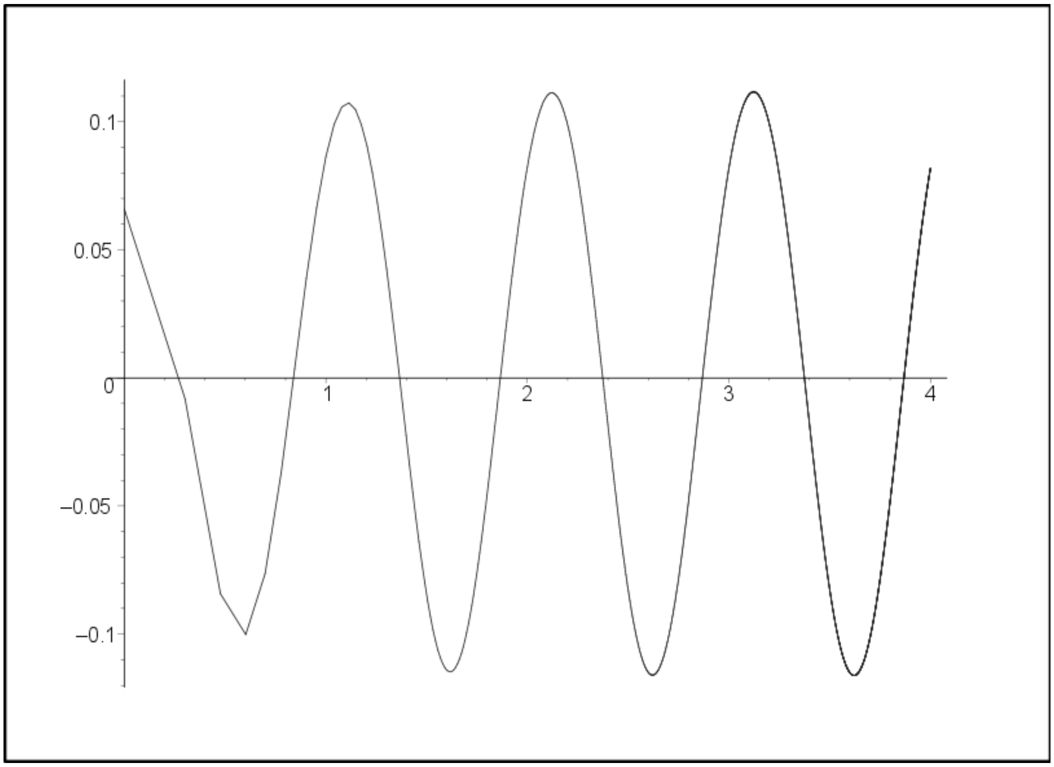
- **Exponential Distribution**

$$f(z,\nu) = \frac{1}{a} \exp\left(-\frac{z}{a}\right), 0 \leq z.$$

$$\begin{aligned} F &= 1, \beta = 1, \alpha = 0, \\ A_n - iB_n &= \frac{2i}{\ln(10)} \Gamma(Y\imath), \\ {A_n}^2 + {B_n}^2 &= \frac{2}{n\ln(10)\sinh(\pi Y)}. \\ Y &= \frac{2\pi n}{\ln(10)}. \end{aligned} \tag{60}$$

$$\begin{aligned} &A_1 = .01308, B_1 = -.01256, \\ &A_2 = -.00002436, B_2 = .000175. \end{aligned} \tag{61}$$

$$\begin{aligned} \rho(k,a,n) &= \log(k+1) - \log(k) \\ &+ \sum_{j=1}^{n} A_j(\sin(js_1) - \sin(js)) \\ &+ \sum_{j=1}^{n} B_j(\cos(js_1) - \cos(js)), \\ s_1 &= 2\pi\frac{\log(k+1)}{a}, s = 2\pi\frac{\log(k)}{a}. \end{aligned}$$



$\ln(10)k(\rho(k) - \log(k+1) + \log(k))$ vs s=$\log(k)$

Figure 3: Exponential Distribution

- **Gamma Distribution**

$$f(z, a, \delta) = \frac{1}{a}\left(\frac{z}{a}\right)^{\delta-1} \exp\left(-\frac{z}{a}\right)\Gamma(\delta)^{-1}, 0 \leq z.$$

$F = 1, \alpha = \delta - 1, \beta = 1, 0 < \delta,$

$$A_n(\delta) - iB_n(\delta) = \frac{1}{\pi n}\Gamma(\delta + Yi)\Gamma(\delta)^{-1}, \tag{62}$$

$$A_n(\delta)^2 + B_n(\delta)^2 = \frac{1}{n^2\pi^2}\Gamma(\delta + Yi)\Gamma(\delta - Yi)\Gamma(\delta)^{-2}, . \tag{63}$$

$$Y = \frac{2\pi n}{\ln(10)}, \tag{64}$$

$$s_1 = 2\pi \log\left(\frac{k+1}{a}\right), s = 2\pi \log\left(\frac{k}{a}\right). \tag{65}$$

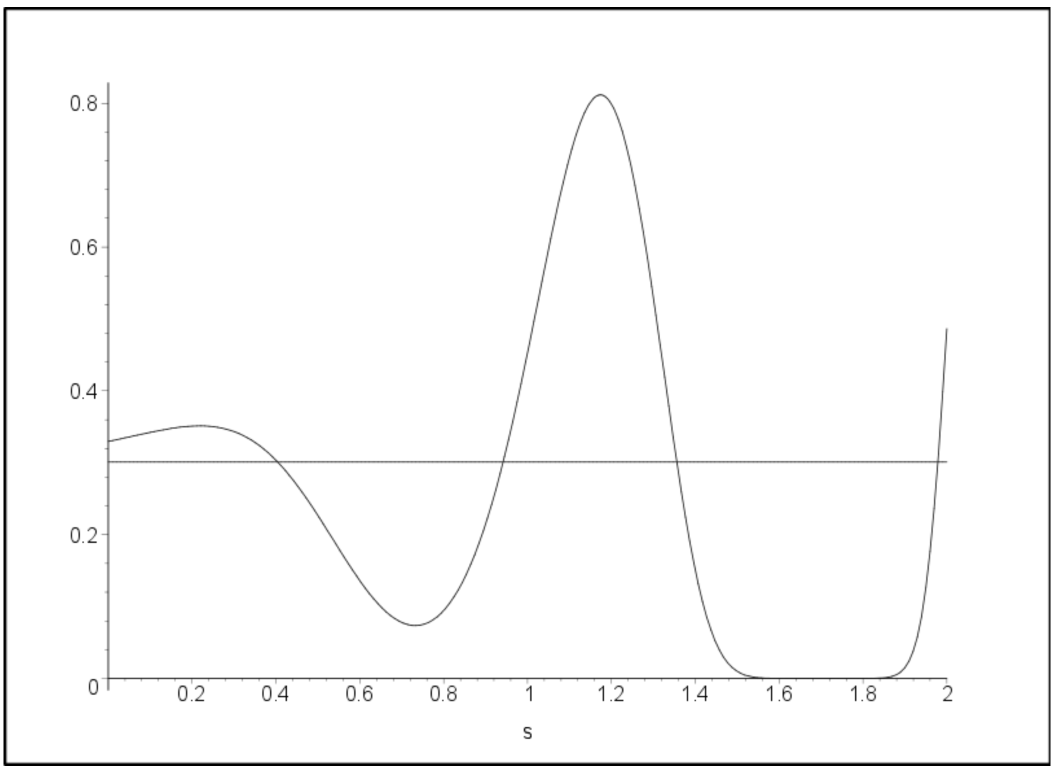
$$A_1(2) = -.02117, B_1(2) = -.04829, \tag{66}$$
$$A_2(2) = .00093, B_2(2) = .0003105. \tag{67}$$

We have a non-monotonic behavior of $\rho(k, a, \delta)$ as a function of δ

δ	1	2	3	4	5	6	7	8	9
$\rho(1,1,\delta)$	.32960	.3432	.2468	.1342	.0782	.0829	.1342	.2205	.3309
$\rho(9,1,\delta)$	.048923	.038394	.020912	.016467	.027004	.04886	.07675	.10370	.12285

Table 6: Table Caption 6

We can prove that as $\delta \to +\infty$, the function $\rho(k, a, \delta)$ has 1-periodic behavior in $\log(\delta)$.



$\rho(k = 1, a = 1, \delta)/\log(2)$ vs $s = \log(\delta)$

Figure 4: Gamma Distributio

Similar analyses are possible for other distributions, like the χ^2-distribution, Student's t-distribution, the Fisher F-distribution, and others.

- **Normal Distribution** with $\mu = 0$

$$f(z,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right), 0 \le z.$$

$$F = \frac{1}{2}, a = \sigma\sqrt{2}, \alpha = 0, \beta = 2, \tag{68}$$

$$A_n - iB_n = \frac{F}{\pi n}\Gamma\left(\frac{1}{2} + Y\imath\right)\Gamma\left(\frac{1}{2}\right)^{-1}, Y = \frac{\pi n}{\ln(10)}, \tag{69}$$

$$\Lambda(n)^2 = \left(4\pi^2 n^2 \cosh\left(\frac{\pi^2 n}{\ln(10)}\right)\right)^{-1},$$

$$s_1 = 2\pi \log\left(\frac{k+1}{\sigma\sqrt{2}}\right), s = 2\pi \log\left(\frac{k}{\sigma\sqrt{2}}\right).$$

$$\rho(k,\sigma,n) = \frac{1}{2}\log\left(1 + \frac{1}{k}\right) + \sum_{j=1}^{n} A_n \left(\sin\left(js_1\right) - \sin\left(js\right)\right)$$
$$+ \sum_{j=1}^{n} B_n \left(\cos\left(js_1\right) - \cos\left(js\right)\right).$$

From this we have that $\rho(k,\mu)$ is a periodic function of $\log(\sigma)$ with period 1.

$$\Lambda(1) = .02640, A_1 = .01621, B_1 = .02081,$$
$$\Lambda(2) = 1.548 \times 10^{-3}, A_2 = 1,546 \times 10^{-3}, B_2 = -4.002 \times 10^{-5},$$
$$\Lambda(3) = 1.210 \times 10^{-4}.$$

We can compare the values of $\rho(k,\sigma,n)$, depending on how many terms of the following are taken into account.

$$\frac{\rho(1,1,0)}{\rho(1,1,\infty)} - 1 = -.1627, \frac{\rho(1,1,1)}{\rho(1,1,\infty)} - 1 = -1.634\times10^{-2}, \frac{\rho(1,1,2)}{\rho(1,1,\infty)} - 1 = 1.202\times10^{-6}$$

7.4 A Second Family of Distributions

Another class of distributions [14] is

$$f(z) = \begin{cases} 0, & z < 0 \\ F a^{(-1)} \beta \left(\frac{z}{a}\right)^{\alpha} \left(1 + \left(\frac{z}{a}\right)^{\beta}\right)^{-\delta} \Gamma(\delta)\,\Gamma\left(\delta - \frac{\alpha}{\beta} - \frac{1}{\beta}\right)^{-1} & 0 \le z, 0 < a, 0 < \alpha, \\ \Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta}\right)^{-1}, & 0 < \beta, 0 < \delta - \frac{\alpha}{\beta} - \frac{1}{\beta}. \end{cases}$$

$$\begin{aligned}
\rho(k,a,\alpha,\beta,\delta) = \digamma\, &(\log(k+1) - \log(k)) + \\
&\sum_{n=1}^{\infty} A_n(\alpha,\beta,\delta)\left(\sin\left(ns_1\right) - \sin\left(ns\right)\right) + \\
&\sum_{n=1}^{\infty} B_n(\alpha,\beta,\delta)\left(\cos\left(ns_1\right) - \cos\left(ns\right)\right), \\
s_1 = 2\pi n \log&\left(\frac{k+1}{a}\right), s = 2\pi n \log\left(\frac{k}{a}\right)
\end{aligned} \tag{70}$$

From this we have that $\rho(k,a,\alpha,\beta,\delta)$ is a periodic function of $\log(a)$ with period 1.

Where

$$\begin{aligned}
A_n(\alpha,\beta,\delta) - iB_n(\alpha,\beta,\delta) = \digamma \frac{1}{\pi n} &\Gamma\left(\delta - \frac{\alpha}{\beta} - \frac{1}{\beta} - Yi\right)\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta} + Yi\right) \\
&\Gamma\left(\delta - \frac{\alpha}{\beta} - \frac{1}{\beta}\right)^{-1}\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta}\right)^{-1}, \\
Y = &\frac{2\pi n}{\beta \ln(10)}, i^2 = -1,
\end{aligned} \tag{71}$$

If

$$\delta = \frac{2(\alpha+1)}{\beta}$$

then

$$\begin{aligned}
A_n(\alpha,\beta) = \digamma \frac{1}{\pi n} &\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta} - Yi\right)\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta} + Yi\right) \\
&\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta}\right)^{-1}\Gamma\left(\frac{\alpha}{\beta} + \frac{1}{\beta}\right)^{-1} \\
B_n(\alpha,\beta) = 0.&
\end{aligned}$$

- **Cauchy Distribution** with $\mu = 0$

$$f(z,a) = \frac{1}{\pi a}\frac{1}{1+\left(\frac{z}{a}\right)^2}, 0 < a.$$

$$\digamma = \frac{1}{2}, \alpha = 0, \beta = 2, \delta = 1$$

$$\begin{aligned}
A_n &= \frac{\digamma}{\pi n}\Gamma\left(\frac{1}{2} - Y\imath\right)\Gamma\left(\frac{1}{2} + Y\imath\right)\Gamma\left(\frac{1}{2}\right)^{-2} \\
&= \digamma\left(\pi n \cosh\left(\frac{\pi^2 n}{\ln(10)}\right)\right)^{-1}, Y = \frac{\pi n}{\ln(10)}, \\
B_n &= 0.
\end{aligned}$$

$$A_1 = .004380, A_2 = .000063012. \tag{72}$$

- **F-Ratio Distribution**

$$f(z) = \left(\frac{\nu}{\omega}\right)^{\frac{\nu}{2}} z^{\frac{\nu}{2}-1}\left(1+\frac{z\nu}{\omega}\right)^{-\delta}\Gamma(\delta)\Gamma\left(\delta-\frac{\alpha}{\beta}-\frac{1}{\beta}\right)^{-1}, \delta=\frac{\nu}{2}+\frac{\omega}{2}, 0<\nu, 0<\omega.$$

$$F=1, a=\frac{\omega}{\nu}, \alpha=\frac{\nu}{2}-1, \beta=1, \delta=\frac{\nu}{2}+\frac{\omega}{2}$$

$$\begin{aligned} A_n(\nu,\omega)-iB_n(\nu,\omega) &= \frac{1}{\pi n}\Gamma\left(\frac{\omega}{2}-Y\imath\right)\Gamma\left(\frac{\nu}{2}+Y\imath\right) \\ &\Gamma\left(\frac{\omega}{2}\right)^{-1}\Gamma\left(\frac{\nu}{2}\right)^{-1}, \\ Y &= \frac{2\pi n}{\ln(10)}, \\ s_1 &= 2\pi\log\left(\frac{\nu(k+1)}{\omega}\right), s=2\pi\log\left(\frac{\nu k}{\omega}\right). \end{aligned} \tag{73}$$

If $\omega=\nu$, we have

$$\begin{aligned} A_n(\nu,\nu) &= \frac{1}{\pi n}\Gamma\left(\frac{\nu}{2}-Yi\right)\Gamma\left(\frac{\nu}{2}+Yi\right)\Gamma\left(\frac{\nu}{2}\right)^{-2}, \\ B_n(\nu,\nu) &= 0, Y=\frac{2\pi n}{\ln(10)}. \end{aligned}$$

$$A_1(3,3) = .003705, A_2(3,3) = .00000136. \tag{74}$$

One example of this type of distribution is the ratio of two positive random numbers ($\alpha=0, \beta=2, \delta=1$).

7.4.1 DSLD for the Ratio of two Positive Numbers

If we take any two positive numbers x and y, then what is the distribution of the leading significant digits of $\frac{y}{x}$?

Introduce new variables R and φ (where $0<\varphi<\frac{\pi}{2}, 0<R$)

$$\begin{aligned} &x = R\cos(\varphi), y=R\sin(\varphi), \\ &\frac{y}{x} = \tan(\varphi). \end{aligned}$$

Then for the leading digit k we have

$$10^m k < \tan(\varphi_k) < 10^m(k+1), 1\le k\le 9, -\infty<m<\infty, m \text{ is integer.}$$

and so

$$\arctan(10^m k) < \varphi_k < \arctan(10^m (k+1)).$$

$$\varphi_k = \sum_{m=-\infty}^{\infty} \left(\arctan(10^m (k+1)) - \arctan(10^m k)\right). \tag{75}$$

$\rho(k)$ of leading digit k is

$$\rho(k) = \frac{2}{\pi}\varphi_k. \tag{76}$$

k	1	2	3	4	5	6	7	8	9
$\rho(k)$	.3092	.1691	.1184	.09361	.07903	.06824	.06023	.05394	.04811

Table 7: Table Caption 7

It is clear that this distribution is connected with the Cauchy distribution $\left(PDF = \frac{1}{\pi}\frac{1}{1+z^2}\right)$

Here we present some areas of $\frac{2}{\pi}\left(\arctan(10^m(k+1)) - \arctan(10^m k)\right)$ for different k and m:

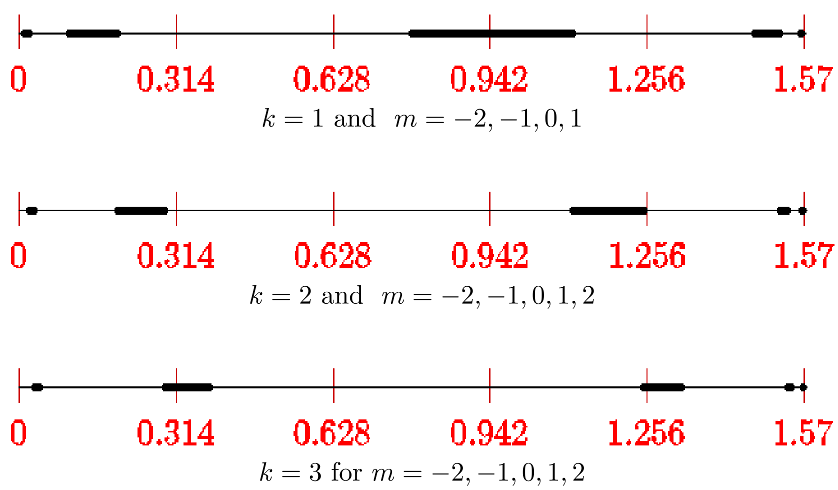
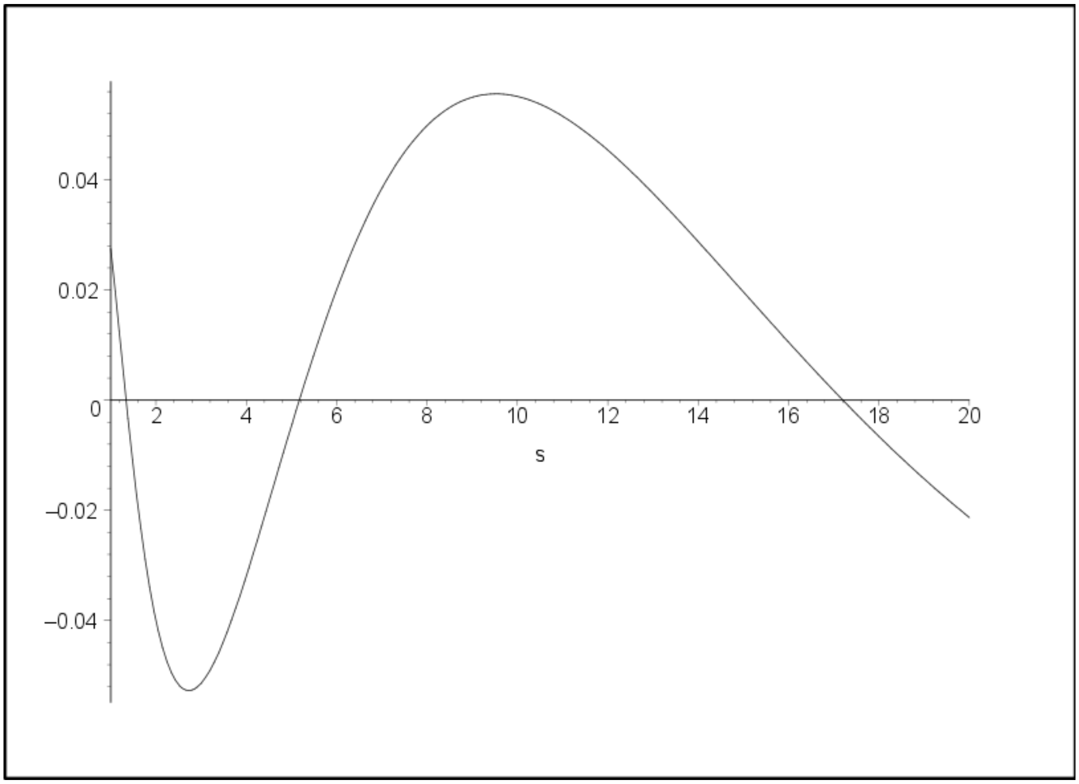


Figure 5: The Ratio of two Positive Numbers

The ratio of the oscillatory part to $\log(1 + 1/k)$ in $\rho(k)$ has the form



$\frac{\rho(k)}{\log(1+1/k)} - 1$ vs s=k

Figure 6: The ratio of the oscillatory to non-oscilattory parts

Then we have

$$\begin{aligned}\rho(k, N) &= \log(k+1) - \log(k) + \\ &\quad \sum_{n=1}^{N} A_n \left(\sin(2\pi n \log(k+1)) - \sin(2\pi n \log(k))\right), \\ A_n &= \left(\pi n \cosh\left(\frac{n\pi^2}{\ln(10)}\right)\right)^{-1}.\end{aligned} \tag{77}$$

If we calculate the difference from the NBL ($k = 1$), we get

$$100\% \left(\frac{0.3092}{0.3010} - 1\right) = 2.75\%.$$

The full solution is $\rho(k, \infty)$. The one-term approximation of this solution has only a small deviation from the exact solution

$$\rho(k,1)=\log(k^{-1}+1)+\left(\pi\cosh\left(\frac{\pi^2}{\ln(10)}\right)\right)^{-1}\left(\sin(2\pi\log(k+1))-\sin(2\pi\log(k))\right),$$

$$\frac{\rho(k,1)}{\rho(k,\infty)}-1=O\left(10^{-4}\right).$$

Now we consider the ratio of two random variables $z=\frac{y}{x}$ selected from normal distributions.

- **Normal Distributions** ($0<\sigma_x, 0<\sigma_y, \mu_x=0, \mu_y=0$) with PDF

$$\begin{aligned}
f(z) &= \int\limits_{x=0}^{\infty} x f_{norm}(xz,0,\sigma_y) f_{norm}(x,0,\sigma_x)dx \\
&= \frac{\sigma_x\sigma_y}{2\pi\left(\sigma_x^2 z^2+\sigma_y^2\right)}, \\
\digamma &= \int\limits_{z=0}^{\infty} f(z)dz = \frac{1}{4}.
\end{aligned}$$

We have 1/4 because we consider only the first quadrant ($0<x, 0<y$).

Then

$$\begin{aligned}
\rho(k,\sigma_x,\sigma_y) &= \digamma\left(\begin{array}{c}\log(1+k)-\log(k)+\\ \sum\limits_{n=1}^{\infty}A_n\left(\sin(ns_1)-\sin(ns)\right)\end{array}\right), \\
A_n &= \left(\pi n\cosh\left(\frac{n\pi^2}{\ln(10)}\right)\right)^{-1}, \\
s_1 &= 2\pi\log\left(\frac{\sigma_x(k+1)}{\sigma_y}\right), s=2\pi\log\left(\frac{\sigma_x k}{\sigma_y}\right).
\end{aligned}$$

$A_1=8.7556\times10^{-3}, A_2=6.0230\times10^{-5}$.

From this we obtain a DSLD close to the NBL, multiplied by $\digamma$.

Now we consider the ratio of two random variables $z=\frac{y}{x}$ selected from exponential distributions.

- **Exponential Distribution** $(0<\lambda_x, 0 < \lambda_y)$

$$
\begin{aligned}
f_x(x) &= \frac{1}{\lambda_x} \exp\left(-\frac{x}{\lambda_x}\right), f_y(y) = \frac{1}{\lambda_y} \exp\left(-\frac{y}{\lambda_y}\right).\\
f(z) &= \int_{x=0}^{\infty} x f_y(xz) f_x(x) dx\\
&= \frac{\lambda_x \lambda_y}{(\lambda_x z + \lambda_y)^2},\\
F &= \int_{z=0}^{\infty} f(z) dz = 1.
\end{aligned}
$$

Then

$$
\begin{aligned}
\rho(k, \lambda_x, \lambda_y) &= \left(\begin{array}{c} \log(1+k) - \log(k) + \\ \sum_{n=1}^{\infty} A_n \left(\sin(ns_1) - \sin(ns)\right) \end{array} \right),\\
A_n &= \frac{2\pi}{\ln(10)} \sinh\left(\frac{n^2\pi}{\ln(10)}\right)^{-1},\\
s_1 &= 2\pi \log\left(\frac{\lambda_x(k+1)}{\lambda_y}\right), s = 2\pi \log\left(\frac{\lambda_x k}{\lambda_y}\right).
\end{aligned}
$$

$$A_1 = 1.0326 \times 10^{-3}, A_2 = 1.9540 \times 10^{-7}.$$

From this we obtain a DSLD close to the NBL.

Now we consider the ratio of two random variables $z = \frac{y}{x}$ selected from LogNormal distributions.

- **Ratio** of two positive numbers with **LogNormal distributions**

$$
\begin{aligned}
z &= \frac{y}{x},\\
f(z) &= \int_{x=0}^{\infty} x f_y(xz, \mu_y, \sigma_y) f_x(x, \mu_x, \sigma_x) dx\\
&= f(z, \mu, \sigma),\\
\mu &= \mu_x - \mu_y, \sigma^2 = \sigma_x^2 + \sigma_y^2,\\
F &= \int_{z=0}^{\infty} f(z) dz = 1.
\end{aligned}
$$

and the PDF for z has a LogNormal form, see (48).

7.5 DSLD for Normal Distribution

Because the normal distribution is common in many applications, let us consider it in more detail.

$$
\begin{aligned}
f(z,\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right),\\
F(z,\mu,\sigma) &= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{z-\mu}{\sigma\sqrt{2}}\right),\\
\digamma &= \int\limits_{z=0}^{\infty} f(z)dz = 1 - F(0,\mu,\sigma)\\
&= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\mu}{\sigma\sqrt{2}}\right).
\end{aligned}
$$

If $\mu \neq 0$, we have $\left(\xi = \frac{\mu}{\sigma\sqrt{2}}\right)$

$$
\begin{aligned}
G(s,\mu,\sigma) &= C\left(\xi\right)s + Q_1\left(s,\xi\right),\\
C\left(\xi\right) &= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\xi\right),\\
Q_1\left(s,\xi\right) &= Q_1\left(s+1,\xi\right),\\
s &= \log(k) - \log(\sigma\sqrt{2}), s_1 = \log(k+1) - \log(\sigma\sqrt{2}),\\
\rho(k,\mu,\sigma) &= A\left(\xi\right)\left(\log(k+1) - \log(k)\right) + Q_1\left(s_1,\xi\right) - Q_1\left(s,\xi\right).
\end{aligned}
$$

The function $\rho(k,\mu,\sigma)$ is invariant under the transformations $\mu \to 10^n\mu, \sigma \to 10^n\sigma$ with integer n.

When $\mu = 0$ see 68. In the general case, we will use the convenient form (111)

$$
\rho(k) = \int\limits_{z=0}^{\infty} f(z)\left(\left\lfloor \log\left(\frac{z}{k}\right)\right\rfloor - \left\lfloor \log\left(\frac{z}{k+1}\right)\right\rfloor\right)dz. \tag{78}
$$

In some cases we can find asymptomatic formulas.

- If $0 < \mu, \mu = O(1), \frac{\sigma}{\mu} \ll 1, \digamma = 1$

  It is convenient to make the change

$$
\begin{aligned}
z &= \mu + \sigma\sqrt{2}t,\\
\rho(k) &= \frac{1}{\sqrt{\pi}} \int\limits_{t=-\frac{\mu}{\sigma\sqrt{2}}}^{\infty} \exp\left(-t^2\right)\left(\left\lfloor \log\left(\frac{\mu+\sigma\sqrt{2}t}{k}\right)\right\rfloor - \left\lfloor \log\left(\frac{\mu+\sigma\sqrt{2}t}{k+1}\right)\right\rfloor\right)dt.
\end{aligned}
$$

As $\frac{\sigma}{\mu} \to 0$, we have

$$\rho(k) = \begin{cases} \frac{1}{2} & \log\left(\frac{\mu}{k}\right) - \left\lfloor \log\left(\frac{\mu}{k}\right) \right\rfloor = 0, \\ \frac{1}{2} & \log\left(\frac{\mu}{k+1}\right) - \left\lfloor \log\left(\frac{\mu}{k+1}\right) \right\rfloor = 0 \\ 1 & \left\lfloor \log\left(\frac{\mu}{k}\right) \right\rfloor - \left\lfloor \log\left(\frac{\mu}{k+1}\right) \right\rfloor > 0 \\ 0 & \text{other.} \end{cases} \tag{79}$$

$$\Omega(s) = \begin{cases} -\frac{1}{2}, & s = 0, \\ \frac{1}{2}, & s = 1, \\ 1, & 1 < s, \\ 0, & \text{other.} \end{cases},$$

$$\Omega(1) - \Omega(0) = 1.$$

$$\rho(k) = \Omega(s_1) - \Omega(s),$$
$$s = \log\left(\frac{k}{\mu}\right) - \left\lfloor \log\left(\frac{k}{\mu}\right) \right\rfloor,$$
$$s_1 = \log\left(\frac{k+1}{\mu}\right) - \left\lfloor \log\left(\frac{k}{\mu}\right) \right\rfloor.$$

- In the general case, the formulas for A_n and B_n too bulky and it is better to use a numerical method for calculations.

Let find the distribution of the leading digits for the deviation from μ, $z = x - \mu$

$$F = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\mu\sqrt{2}}{2\sigma}\right).$$

$$A_n - \imath B_n = \frac{1}{n\pi^{\frac{3}{2}}} \int_{z=0}^{\infty} x^{iY} \exp\left(-(x-\xi)^2\right) dx,$$
$$Y = \frac{2n\pi}{\ln(10)}.$$

As an example, we consider the set of parameters as in [15] $\mu = 1.1, \sigma = 0.25$ then

$$F = \frac{1}{2} + \frac{1}{2} \operatorname{erf}(3.1113) = .99999.$$

k	1	2	3	4	5	6	7	8	9
$\rho(n=\infty)$	.6550	.0010	.0200	.0050	.01500	.0320	.0600	.0970	.1330
$\rho(n=4)$	.6540	.0021	.0240	.0046	.01372	.0347	.0583	.0952	.1348

Table 8: Table Caption 8

The structure of the DSLD depends on the parameters of the underlying distribution, $\mu = 1.1$ and $\sigma = 0.25$ in this case, which is equivalent to $\mu = 11$ and $\sigma = 2.5$, with sample numbers mainly lying in the interval $\mu \pm 3\sigma$ $[3.5, 18.5]$ For $\mu = 100, \sigma = 15$ (which is equivalent to $\mu = 10$, $\sigma = 1.5, [5.5, 14.5]$), we have

k	1	2	3	4	5	6	7	8	9
$\rho(n=\infty)$	.5000	.0000	.0000	.0000	.00380	.01900	.0680	.1620	.2480
$\rho(n=6)$	.4980	-0.0009	.0007	.00183	.00082	.02162	.0663	.1616	.2497

Table 9: Table Caption 9

Although there is a logarithmic term in the representation of the DSLD, the distribution is not monotonous or tending to the NBL.

From (79) we can see that $\rho(k)$ is a 1-periodic function of $\log(k)$ (for k positive) and $\log(\sigma)$

$$\begin{aligned}\rho(k) = {} & \frac{1}{2}\left(\log(k+1) - \log(k)\right) \\ & + Q_1(\log(k+1) - \lfloor\log(k)\rfloor + \lfloor\log(\sigma)\rfloor - \log(\sigma)) \\ & - Q_1(\log(k) - \lfloor\log(k)\rfloor + \lfloor\log(\sigma)\rfloor - \log(\sigma)), \\ & k = 1, 2, \ldots.\end{aligned} \tag{80}$$

k	1	2	3	4	5	6	7	8	9
$\rho(\sigma=1)$	.17987	.06434	.04330	.04050	.03860	.03668	.03446	.03227	.02980

Table 10: Table Caption 10

and

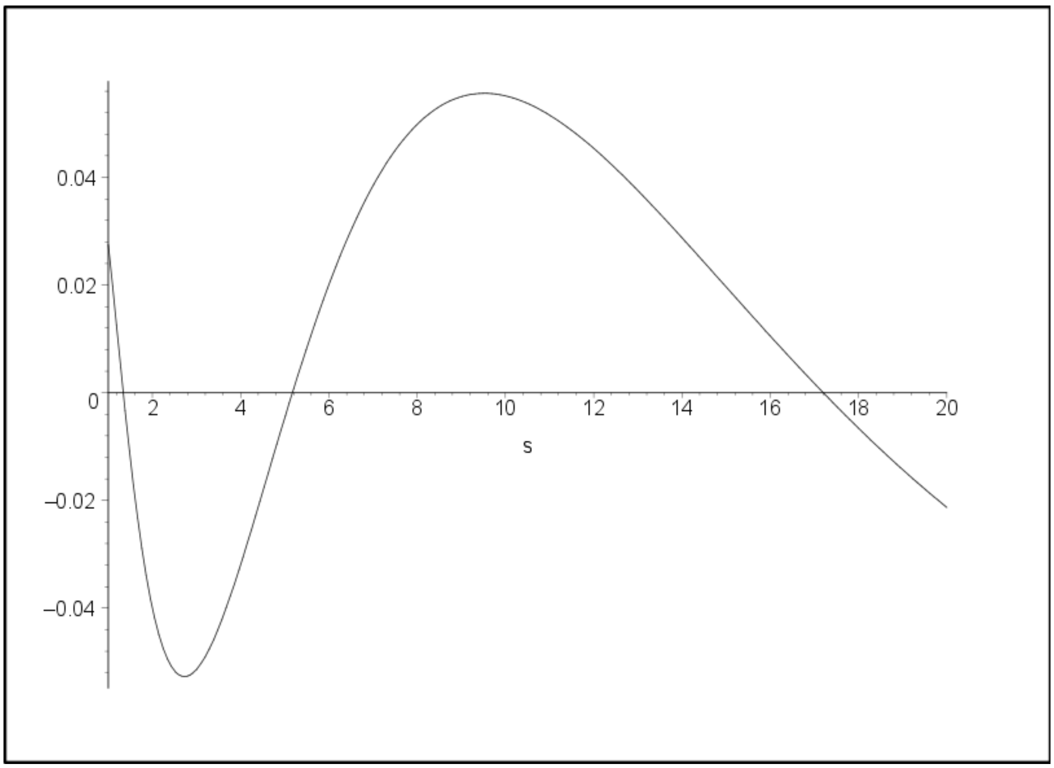
σ	1	2	3	4	5	6	7	8	9
$\rho(k=1)$	.17987	.17200	.13213	.11000	.10770	.11911	.13640	.15400	.16874

Table 11: Table Caption 11

$$G(s) = \frac{s}{2} + 0.026418\sin(2\pi s) - 0.000938\cos(2\pi s)$$
$$- 0.000536\sin(4\pi s) - 0.001459\cos(4\pi s).$$

$$\rho(k) \approx G(\log(k+1) - \lfloor\log(k)\rfloor + \lfloor\log(\sigma)\rfloor - \log(\sigma))$$
$$- G(\log(k) - \lfloor\log(k)\rfloor + \lfloor\log(\sigma)\rfloor - \log(\sigma)),$$
$$k = 1, 2, \ldots.$$

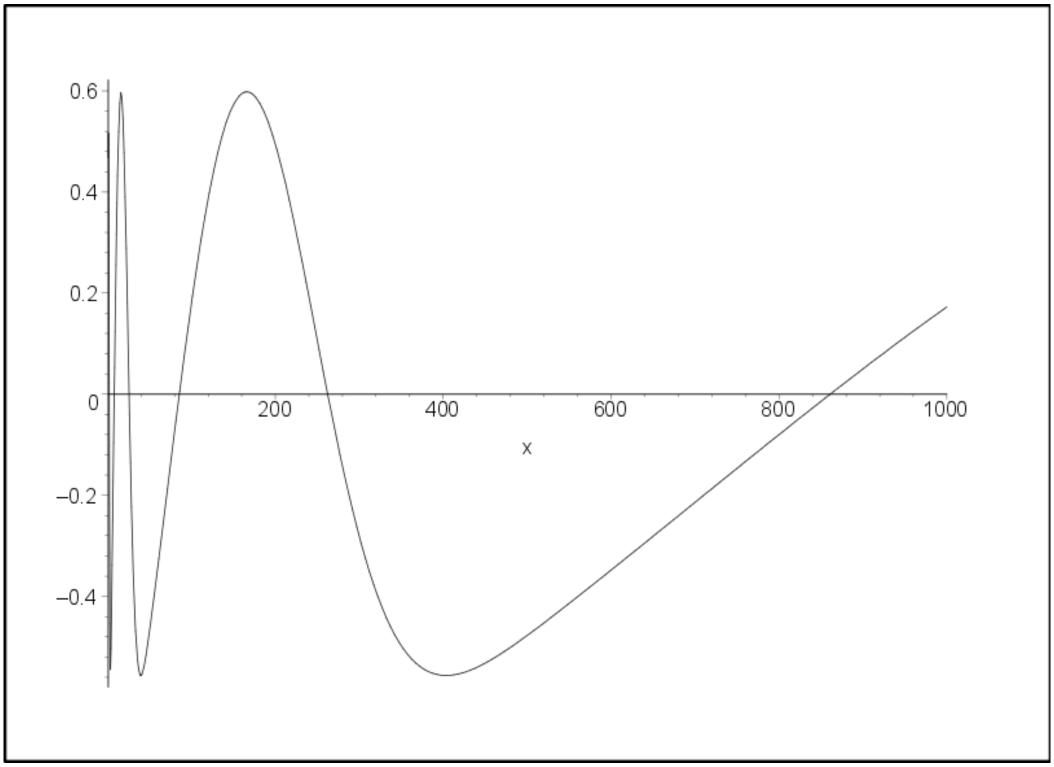
Here are some illustrations of the ratio $\rho(k,\sigma,\mu)/\log(1+1/k)$



$\rho(k=1,\sigma,\mu=0)/\log(2) - 1$ vs $s = \sigma$.

Figure 7: Normal Distribution, $k = 1$ and $\mu = 0$

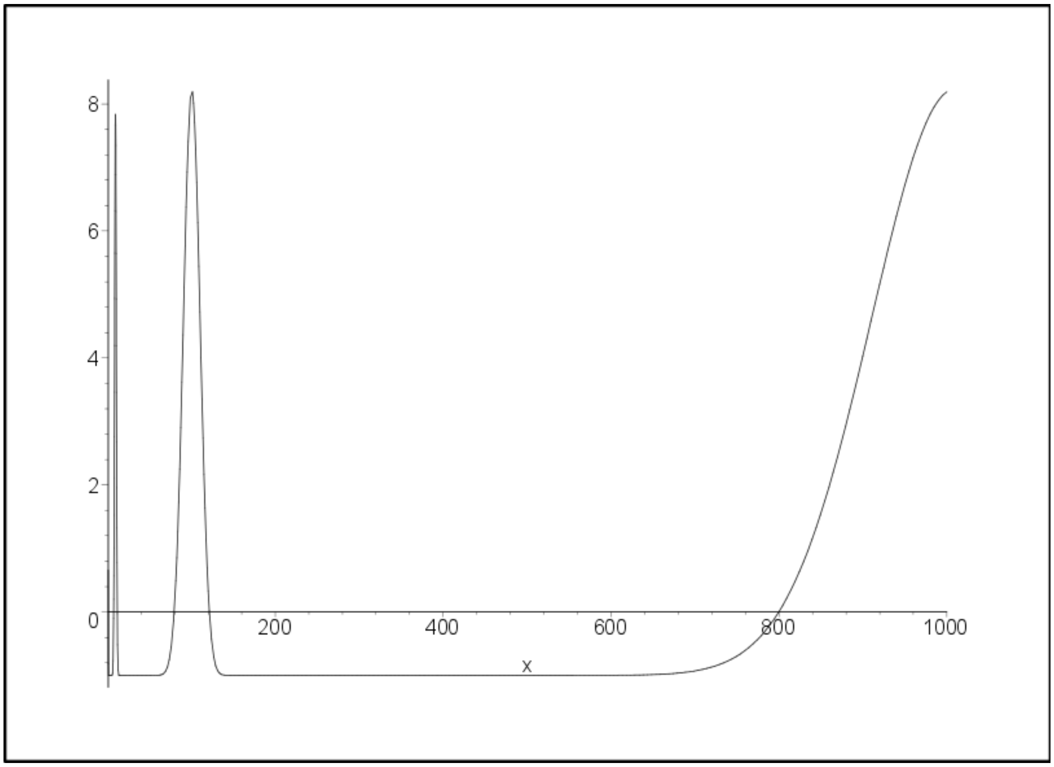
and $(\sigma = 1, \mu = 1)$



$\rho(k, \sigma = 1, \mu = 1)/(F \log(1 + 1/k)) - 1$ vs k

Figure 8: Normal Distribution, $\sigma = 1$ and $\mu = 0$

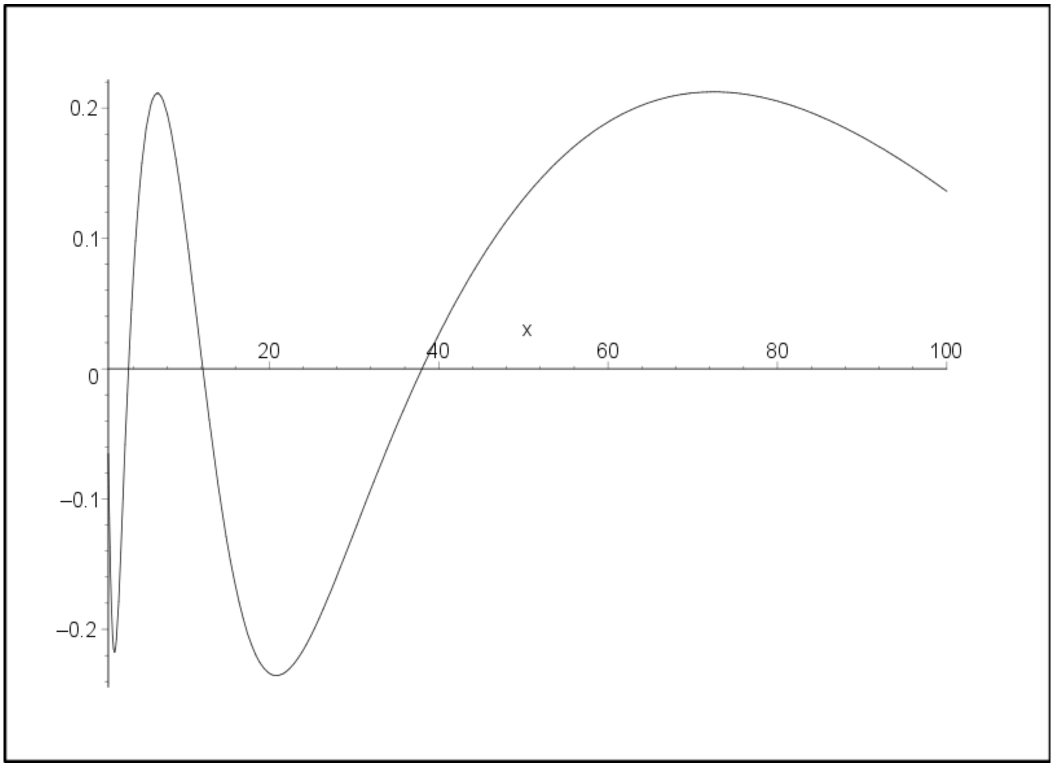
and $(\sigma = 0.1, \mu = 1)$



$\rho(k, \sigma = 0.1, \mu = 1)/(F \log(1 + 1/k)) - 1$ vs k

Figure 9: Normal Distribution, $\sigma = 0.1$ and $\mu = 1$

in this case, the contribution of the oscillating term is up to 8 times greater than that of the logarithmic.



$\rho(k, \sigma = 1, \mu = -1)/(F \log(1 + 1/k)) - 1$ vs k

Figure 10: Normal Distribution, $\sigma = 1$ and $\mu = -1$

An easy kind of approximation has the form

$$\begin{aligned}\rho(k, \sigma) \approx &\frac{1}{2} \log\left(\frac{k+1}{k}\right) \\ &+ \frac{1}{12\pi}\left(\sin\left(2\pi(\log(k+1) - \log(\sigma))\right) - \sin\left(2\pi(\log(k) - \log(\sigma))\right)\right), \\ &k = 1, 2, \ldots.\end{aligned}$$

For negative numbers x, we have

$$10^m(k-1) \le x \le 10^m k$$

$$
\begin{aligned}
\rho(k) = \frac{1}{2}\left(\log(|k|+1) - \log((|k|)\right) + Q_1(\log(|k|+1) - \lfloor\log((|k|)\rfloor + \lfloor\log(\sigma)\rfloor - \log(\sigma)) \quad (81)\\
- Q_1(\log((|k|) - \lfloor\log((|k|)\rfloor + \lfloor\log(\sigma)\rfloor - \log(\sigma)),\\
k = -1, -2, \ldots.
\end{aligned}
$$

It is important to emphasize that in spite of the possible argument about its insignificant value, the role of the functions Q_1 can be substantial. So if we look at the dependence on σ, the difference can reach more than 50%.

$$
\begin{aligned}
\rho(k=1, \sigma=1.327, \mu=0) = .19287,\\
\rho(k=1, \sigma=4.613, \mu=0) = .10657,\\
\frac{\rho(1, 1.327) + \rho(1, 4.613)}{2} = .14972,\\
100\%\frac{2(\rho(1, 1.327) - \rho(1, 4.613))}{\rho(1, 1.327) + \rho(1, 4.613)} = 57.6\%.
\end{aligned}
$$

When $\mu \neq 0$, we have

$$
\begin{aligned}
G(s, \mu, \sigma) &= A\left(\frac{\mu}{\sigma}\right)s + Q_1\left(s, \frac{\mu}{\sigma}\right),\\
A\left(\frac{\mu}{\sigma}\right) &= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\mu\sqrt{2}}{2\sigma}\right),\\
Q_1\left(s, \frac{\mu}{\sigma}\right) &= Q_1\left(s+1, \frac{\mu}{\sigma}\right),\\
s &= \log(k) - \log(\sigma), s_1 = \log(k+1) - \log(\sigma),\\
\rho(k, \sigma, \mu) &= A\left(\frac{\mu}{\sigma}\right)(\log(k+1) - \log(k)) + Q_1\left(s_1, \frac{\mu}{\sigma}\right) - Q_1\left(s, \frac{\mu}{\sigma}\right).
\end{aligned}
$$

7.6 Truncated Normal Distribution

The truncated normal distribution is often used in different application.

$$
\begin{aligned}
f(z, \sigma, \mu, a, b) = \begin{cases} C\exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right), & a \leq z \leq b \\ 0, & z < a, b < z \end{cases},\\
C = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\left(\operatorname{erf}\left(\frac{b-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a-\mu}{\sigma\sqrt{2}}\right)\right)^{-1}, -\infty \leq a < b \leq \infty.
\end{aligned}
$$

And

$$
\begin{aligned}
CDF &= C\left(\operatorname{erf}\left(\frac{z-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a-\mu}{\sigma\sqrt{2}}\right)\right)\\
F &= C\left(\operatorname{erf}\left(\frac{\gamma_+-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{\gamma_--\mu}{\sigma\sqrt{2}}\right)\right),\\
\gamma_+ &= \max(b, 0), \gamma_- = \max(a, 0).
\end{aligned}
$$

An expression for $\rho(z, \sigma, \mu, a, b)$ can be found from (**??**).

8 Approximation

Property of function $P(k,x)$

$$P(k,x) = P(k, 10^{\delta}x), x \in \mathbb{R}_{0<}, k \in \mathbb{N}, \delta \in \mathbb{Z}.$$

then

$$\int\limits_{x=0}^{\infty} f(x)P(k,x)dx = \int\limits_{x=1}^{10} \psi(y)P(k,y)dy,$$

$$\psi(y) = \sum_{\delta=-\infty}^{\infty} 10^{\delta} f\left(10^{\delta}y\right).$$

other property of $P(k,x)$

$$P(k,x) = P\left(\frac{k}{10^{\delta}}, x\right), x \in \mathbb{R}_{0<}, k \in \mathbb{N}, \delta \in \mathbb{Z}.$$

then for example

$$P(11,x) = P\left(\frac{11}{10}, x\right).$$

We can consider function $\rho(k)$ on interval $[1,10)$ with discret points $1, \frac{101}{100}, \frac{102}{100}, \ldots, \frac{11}{10}, \frac{12}{10}, \ldots, \frac{19}{10}, \ldots$

$$\rho(k) = \sum_{i=0}^{10^{L}-1} \rho(10^{L+1}k + i), L = \lfloor \log(k) \rfloor.$$

Example

$$k = \frac{11}{10}, L = 0, \rho\left(\frac{11}{10}\right) = \rho(11).$$

Other presentation os s and s_1

$$s = \log(k) - \lfloor \log(k) \rfloor, s_1 = \log(k+1) - \lfloor \log(k) \rfloor,$$

$$s = \log(k) + \lfloor -\log(k) \rfloor, s_1 = \log(k+1) + \lfloor -\log(k+1) \rfloor.$$

8.1 Propertys of $\lfloor x \rfloor$

Because

$$\lfloor -x \rfloor = -\lfloor x \rfloor + \begin{cases} 0, & x \in \mathbb{Z}, \\ -1, & x \notin \mathbb{Z}, \end{cases}$$

Then

$$\lfloor -\log(k+1) \rfloor = -\lfloor \log(k) \rfloor - 1,$$

and

$$\lfloor -\log(k+1) \rfloor = \lfloor -\log(k) \rfloor .$$

- Discrete Ω

$$\begin{aligned} -\Omega(\log(k), a) &= \left\lfloor \log\left(\frac{a}{k}\right) \right\rfloor - \log\left(\frac{a}{k}\right) + \log(a) - \log(k), \\ &= -\left(\log\left(\frac{a}{k}\right) - \left\lfloor \log\left(\frac{a}{k}\right) \right\rfloor\right) + \log(a) - \log(k) \end{aligned}$$

$$\Omega(s,a) = s + \frac{1}{N}\sum_{i=1}^{N} (\log(a_i) - s - \lfloor \log(a_i) - s \rfloor) - \frac{1}{N}\sum_{i=1}^{N} \log(a_i).$$

$$\begin{aligned} \{x\} &= x - \lfloor x \rfloor \\ &= \begin{cases} x, & x \in \mathbb{Z}, \\ \sum_{j=1}^{\infty} \frac{1}{\pi j} \sin(2\pi j x), & x \notin \mathbb{Z}. \end{cases} . \end{aligned}$$

We set the non-essential additive constant to zero.

$$\Omega(s,a) = s + \frac{1}{N}\sum_{i=1}^{N} \{\log(a_i) - s\} + const.$$

$$\frac{1}{N}\sum_{i=1}^{N} \{\log(a_i) - s\} = \sum_{j=1}^{\infty} \Lambda_j \frac{\sin(2\pi j s + \varphi_j)}{\pi j},$$

$$\Lambda_j \exp(\imath \varphi_j) = \frac{1}{\pi j N}\sum_{i=1}^{N} \exp(-2\pi j \imath \{\log(a_i)\}).$$

- Continues Ω

$$\begin{aligned} -\Omega(\log(k), a) &= \int_{z=0}^{\infty} f(z) \left\lfloor \log\left(\frac{z}{k}\right) \right\rfloor dz \\ &= \int_{z=0}^{\infty} f(z) \left(\left\lfloor \log\left(\frac{z}{k}\right) \right\rfloor - \log\left(\frac{z}{k}\right) + \log(z) - \log(k)\right) dz, \\ &= -\int_{z=0}^{\infty} f(z) \left\{\log\left(\frac{z}{k}\right)\right\} dz - \log(k) \int_{z=0}^{\infty} f(z) dz. \end{aligned}$$

$$\Omega(s) = s\int\limits_{z=0}^{\infty} f(z)dz. + \int\limits_{z=0}^{\infty} f(z)\left(\log(z) - s - \lfloor \log(z) - s \rfloor\right) dz + const.$$

We set the non-essential additive constant to zero.

$$\Omega(s) = s\int\limits_{z=0}^{\infty} f(z)dz. + \int\limits_{z=0}^{\infty} f(z)\left\{\log(z) - s\right\} dz.$$

$$\int\limits_{z=0}^{\infty} f(z)\left\{\log(z) - s\right\} dz = \sum_{j=1}^{\infty} \Lambda_j \sin(2\pi j s + \varphi_j),$$

$$\begin{aligned}\Lambda_j \exp(\imath\varphi_j) &= \frac{1}{\pi j}\int\limits_{z=0}^{\infty} f(z)\exp(-2\pi j\imath\left\{\log(z)\right\})dz \\ &= \frac{1}{\pi j}\int\limits_{z=0}^{\infty} f(z) z^{\left(-\frac{2\pi j\imath}{\log(10)}\right)} dz, j = 1, 2, \ldots.\end{aligned}$$

- Connection with van der Corput's methods.
- From (42)

$$\Lambda_n \exp(\varphi_n \imath) = \frac{1}{2\pi n}\left(\int\limits_{z=0}^{1} z^{(-Y\imath)}dz + \int\limits_{z=1}^{\infty} z^{(-2-Y\imath)}dz\right), Y = \frac{2\pi n}{\ln(10)}, n = 1, 2, \ldots$$

$$\Lambda_n = \frac{1}{\pi n}\frac{1}{(1+Y^2)}, \varphi_n = 0.$$

$$\begin{aligned}&\frac{10^{-L}}{18} - 10^{L}\frac{5}{9}\left(\frac{1}{k+1}\frac{1}{k}\right) \\ &\approx \log\left(1+\frac{1}{k}\right) + 0.0377 \times \left(\sin\left(2\pi\log(1+k)\right) - \sin\left(2\pi\log(k)\right)\right), \\ &L = \lfloor \log(k) \rfloor.\end{aligned} \tag{82}$$

$k = 1, 0.3333 \approx 0.3368,$
$k = 15, 0.02867 \approx 0.03047.$

- From (47)

$$\Lambda_n \exp(\varphi_n \imath) = \frac{1}{\pi n} \int\limits_{z=0}^{1} - z^{(-Y\imath)} \ln(z) dz, Y = \frac{2\pi n}{\ln(10)}, n = 1, 2, \ldots$$

$$\Lambda_n = \frac{1}{\pi n} \frac{1}{(1+Y^2)}, \varphi_n = \arctan\left(\frac{2Y}{1-Y^2}\right).$$

$$\frac{10^{-L}}{9}\left(k\ln(k) - (k+1)\ln(k+1) + 1 + \frac{(9L+10)}{9}\ln(10)\right) \tag{83}$$
$$\approx \log\left(1+\frac{1}{k}\right) + 0.0377 \times \left(\sin\left(2\pi\log(1+k) + 2.4391\right) - \sin\left(2\pi\log(k) + 2.4391\right)\right).$$

$k = 1, 0.24135 \approx 0.24171,$
$k = 15, 0.02353 \approx 0.0222.$

- Estimation coefficients in $n!$
- Approximation of function $P(k, x, n)$

$$P(k,x,n) = \left\{\begin{array}{cl} s - s_1, & s \in \mathbb{Z}, s_1 \in \mathbb{Z}, \\ s - s_1 + \frac{1}{2} - \sum\limits_{j=1}^{n} \frac{\sin(2\pi j s_1)}{\pi j}, & s \in \mathbb{Z}, s_1 \notin \mathbb{Z}, \\ s - s_1 - \frac{1}{2} + \sum\limits_{j=1}^{n} \frac{\sin(2\pi j s)}{\pi j}, & s \notin \mathbb{Z}, s_1 \in \mathbb{Z}, \\ s - s_1 + \sum\limits_{j=1}^{n} \frac{\sin(2\pi j s) - \sin(2\pi j s_1)}{\pi j}, & s \notin \mathbb{Z}, s_1 \notin \mathbb{Z}. \end{array}\right\} \tag{84}$$

$$s = \log\left(\frac{x}{k}\right), s_1 = \log\left(\frac{x}{k+1}\right).$$

- $$z = y + x,$$

  where $y \approx N(\mu, \sigma)$ and $x \approx U(0, \varepsilon)$ then

  $$f(z, \mu, \sigma, \varepsilon) = \frac{1}{2\varepsilon}\left(\operatorname{erf}\left(\frac{z-\mu}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{z-\mu-\varepsilon}{\sigma\sqrt{2}}\right)\right).$$

The function $\rho(k)$ weakly depends on relatively small positive changes in the values of the set. This robustness makes it possible to impose additional small perturbations and to construct an approximate formula for DSLD.

- Vector of coefficiens A_i and B_i is w witn $2n+1 (0 \le n)$ components

$$w = [A_0, A_1, B_1, A_2, B_2, \ldots]^\top .$$

and G with $m (0 \le m)$ is vector of experemental values of some $\rho(k)$

$$G = [G_1, G_2, G_3, G_4, G_5, \ldots]^\top .$$

$$M_{i,1} = \log\left(1 + \frac{1}{k}\right), 1 \le i \le m,$$
$$M_{i,j} = \sin\left(2\pi j \log(1+k)\right) - \sin\left(2\pi j \log(k)\right), j \text{ even}, 2 \le j,$$
$$M_{i,j} = \cos\left(2\pi j \log(1+k)\right) - \cos\left(2\pi j \log(k)\right), j \text{ odd}, 3 \le j.$$

Then estimation of vector w is

$$w = \left(M^\top . M\right)^{-1} . M^\top . G.$$

8.2 Discrete Distributions

$$\rho(k) = \log\left(1 + \frac{1}{k}\right) + \sum_{j=1}^{m} U(k, a, j), m \in \mathbb{Z}^*,$$
$$U(k, a, j) = -\frac{1}{\pi j N} \sum_{i=1}^{N} \left(\sin\left(2\pi j \log\left(\frac{a_i}{1+k}\right)\right) - \sin\left(2\pi j \log\left(\frac{a_i}{k}\right)\right)\right).$$

$$s = 2\pi \log(k), s_1 = 2\pi \log(1+k),$$
$$\begin{aligned} S_j(N) &= \frac{1}{\pi j N} \sum_{i=1}^{N} (a_i)^{-\frac{2\pi j}{\ln(10)} \imath} \\ &= \frac{1}{\pi j N} \sum_{i=1}^{N} \exp\left(-2\pi j \imath \log(a_i)\right) \\ &= \frac{1}{\pi j N} \sum_{i=1}^{N} \exp\left(-2\pi j \imath \left\{\log(a_i)\right\}\right). \end{aligned}$$
$$S_j(N) = R \exp(\imath \varphi),$$

Different presentations of $U(k, a, j)$

$$\begin{aligned} U(k, a, j) = &\operatorname{Re}(S_j(N)) \left(\sin(j s_1) - \sin(j s)\right) \\ &+ \operatorname{Im}(S_j(N)) \left(\cos(j s_1) - \cos(j s)\right). \end{aligned}$$

$$U(k, a, j) = R\left(\sin(j s_1 + \varphi) - \sin(j s + \varphi)\right).$$

$$\sum_i (\sin(A_i - s) - \sin(A_i - s_1))$$
$$= \sum_i (\cos(A_i)(\sin(s_1) - \sin(s)) - \sin(A_i)(\cos(s_1) - \cos(s)))$$

$$\left(\sum_i \exp(-\imath A_i)\right) = R\exp(\imath\varphi),$$
$$\sum_i \cos(A_i) = \mathrm{Re}\left(\sum_i \exp(-\imath A_i)\right),$$
$$-\sum_i \sin(A_i) = \mathrm{Im}\left(\sum_i \exp(-\imath A_i)\right).$$

$$U(k,a,j) = \frac{R}{\pi j N}\left(\sin\left(2\pi j \log\left(1+k\right) + \varphi\right) - \sin\left(2\pi j \log\left(k\right) + \varphi\right)\right).$$

8.3 The Fast Fourier Transform.

$$\int\limits_{z=0}^{\infty} f(z)dz = \ln(10)\int\limits_{t=-\infty}^{\infty} g(t)dt,$$
$$\frac{1}{\pi n}\int\limits_{z=0}^{\infty} \exp(-2\pi n\log(z)\imath)f(z)dz = \frac{1}{\pi n}\int\limits_{t=-\infty}^{\infty} \exp(-2\pi n t\imath)g(t)dt,$$
$$g(t) = \ln(10)10^t f(10^t).$$

Forward Discrete Fourier Transform (DFT):

$$X_n = \sum_{j=0}^{N-1}\exp\left(-2\pi n\frac{j}{N}\imath\right)x_j.$$

Inverse Discrete Fourier Transform (IDFT):

$$x_n = \frac{1}{N}\sum_{j=0}^{N-1}\exp\left(2\pi n\frac{j}{N}\imath\right)X_j.$$

8.3.1 Steler's Law

"...George Stigler, a future Nobel Laureate in Economics, claimed that the specific mixture

of uniform distributions with non-uniformly distributed maximum values is an inconsistency. This observation led Stigler (1945) to propose an alternative FSD distribution that was less skewed toward the lower

digits and was derived without the use of such assumptions.

8.3.2 Stigler's FSD Concept

Stigler (1945) reviewed the Newcomb-Benford FSD phenomenon and proposed that the average frequency

of d as a leading significant digit is

$$F(k) = \frac{1}{9}\left(1 + \frac{10}{9}\ln(10) + k\ln(k) - (k+1)\ln(k+1)\right)$$

He arrived at this conclusion by first assuming that the largest entry in the given statistical table is equally

likely to begin with$k = 1, 2, ..., 9$, and that all other entries in the table are randomly selected from the

uniform distribution of numbers smaller than the largest entry. Defining the rth cycle of numbers as being

the interval $[10^r, 10^{r+1}]$ for some real number r, Stigler finds the distribution of FSDs for the highest entry

in a cycle of numbers from the table and then averages the probabilities over all highest entries. Since

table entries are from a uniform distribution, any digit d should have, at the end of the $(r-1)$st cycle,

occurred (10r - 1)/9 times as an FSD out of 10r - 1 numbers, approximately 10r/9 and 10r

, respectively.

For example, at the end of the first cycle, i.e., [10,100), the digit "2" has occurred as an FSD (102 - 1)/9 = 11

2 STIGLER'S FSD CONCEPT 3

times out of 102 - 1 = 99 numbers, including those from all previous cycles.

Title:

Stigler's approach to recovering the distribution of first significant digits in natural data sets

Author:

Lee, Joanne, University of California, Berkeley

Cho, Wendy K. Tam, University of Illinois, Urbana Champaign

Judge, George G, University of California, Berkeley and Giannini Foundation

Publication Date:

01-19-2009

Series:

CUDARE Working Papers

Permalink:

http://escholarship.org/uc/item/9745m98d

Keywords:

Benford's Law, Stigler's Law, Power Law, Maximum Entropy, Distance Measures

..."

8.4 Van der Corput's method

8.5 Weyl Sum

$$C_N(j) = \frac{1}{\pi j N} \sum_{i=1}^{N} \exp(-2\pi j \imath \log(a_i)).$$

Becouse j is integer this is equivalent to

$$C_N(j) = \frac{1}{\pi j N} \sum_{i=1}^{N} \exp(-2\pi j \imath \{\log(a_i)\}).,$$
$$0 \leq \{x\} = x - \lfloor x \rfloor.$$

Let's estimate the magnitude of the modulus of this function (first Weyl observed that)

$$R_N(j)^2 = C_N(j)\bar{C}_N(j)$$

$$R_N(j)^2 = \frac{1}{(\pi j N)^2} \left(N + 2 \sum_{i=1}^{m} \sum_{m=2}^{N} \cos\left(2\pi j \log\left(\frac{a_i}{a_m} \right) \right) \right).2 \leq N.$$

It is of interest to ask if the function $R_N(j)$ can go to $0, j \in \mathbb{N}$,

$$N + 2 \sum_{i=1}^{m-1} \sum_{m=2}^{N} \cos\left(2\pi j \log\left(\frac{a_i}{a_m} \right) \right) = 0.2 \leq N, j \in \mathbb{N}.$$

Weyl, H. "Über ein Problem aus dem Gebiete der diophantischen Approximationen." Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl., 234-244, 1914. Reprinted in Gesammelte Abhandlungen, Band I. Berlin: Springer-Verlag, pp. 487-497, 1968.

Weyl, H. "Über die Gleichverteilung von Zahlen mod. Eins." Math. Ann. 77, 313-352, 1916. Reprinted in Gesammelte Abhandlungen, Band I. Berlin: Springer-Verlag, pp. 563-599, 1968. Also reprinted in Selecta Hermann Weyl. Basel, Switzerland: Birkhäuser, pp. 111-147, 1956.

9 Discussion

This article develops an approach to the manifestation of a certain type of DSLD. We have shown that the DSLD does not have an arbitrary form; rather, it takes one of three forms. In some cases the DSLD can be expressed in closed form or approximated with a high degree of accuracy. We propose an integral representation of the DSLD convenient for analytical and numerical analysis.

The form of an DSLD depends not only on the type of underlying probability distribution, but also on the parameters of this distribution. So when we change the parameters, the form of the DSLD changes.

We stressed at the outset of our discussion the universality of the NBL, that the introduction of this distribution attracted the attention of researchers to this type of problem.

As we can see in Sections 6.1.2 and 6.1.3 the values for the frequency distributions are close to the NBL, but as seen from the explicit formulas, definitely not the NBL.

The literature on NBL has adopted a very strong and, at the same time, rather restrictive principle, namely, to consider as insignificant deviations that are not described in the law. This principle, strictly followed by many researchers, has led to the impression of the universality of this law, but there are other opinions [12], [13], [16]. The NBL has an empirical nature, and in many cases has no procedure for an estimation of the error measurements of experimental data.

Many researchers have drawn conclusions on distributions based on the analysis of a relatively small number of samples, without the evaluation of approximation errors. It should also be noted here that DSLD necessarily depends on properties of both the underlying distribution and the area interval.

These two factors constitute an essential and intrinsic part of the structure, and neither can be neglected when constructing an DSLD. So they are, in this sense, inseparable, and must always appear together. The contributions of the log term and the oscillating term depend on the underlying probability distribution. In some cases, the contribution of the oscillating term may be relatively small, or it can be overwhelming. In many cases, the question arises, what is the measure of the accuracy of the approximations. Sometimes it is rather arbitrary, and may lead to the 'universality' of the NBL. As we have seen from the analysis of the normal distribution (with $0 < \mu$), the presence of a log term does not lead to a distribution resembling the NBL.

We will consider two simple distributions to illustrate the variation of the DSLD depending on the parameters. The behavior of the DSLD is a function of the properties of the underlying distribution and the values of some parameters.

1. If we are risen endless area for sampling, we have, in general, a solution as the sum of the difference of the logarithms (NBL's style, perhaps with an extra multiplicative factor) and a difference of periodic functions (9,11). Moreover, the presence of periodic functions is important because they make up $30\% - 40\%$ of the value of the DSLD. The natural emergence of periodic functions is defined by a transformation group and infinite boundaries.

2. The presence of finite boundaries leads mainly to distributions of the third type (10)

$$\rho(k) = \Omega(\log(k+1) - \lfloor \log(k) \rfloor) - \Omega(\log(k) - \lfloor \log(k) \rfloor). \tag{85}$$

As shown above in the case of the production of n numbers obeying a UD, with an increasing number of factors, the main contribution gives the members in the inner region and the edge interval effectively does not play a significant role. But even small boundary effects eliminate the presence of a periodic function. These types of underlying distributions are perhaps the best candidates for the NBL.

Knowing the possible structure of the function $\rho(k)$, we can offer a procedure to make a Monte Carlo approximation with experimental data. Because with $1 \ll \log(k))$ we have from (101)

$$\rho(k) = \frac{1}{\ln(10)10^s}\left(A + \frac{dQ_1(s)}{ds}\right) + o\left(10^{-s}\right), s = \log(k).$$

For example, for (9,11), we can get from $1 \ll \log(k)$ (in many cases $3 \leq \log(k)$) after statistical experimental density function $\rho(1 \ll \log(k)) = \eta(s)), s = \log(k)$.
Then make the approximation

$$10^s \ln(10) f(s) \approx a_0 + a_1 \sin(2\pi s) + b_1 \cos(2\pi s) + \ldots, S \leq s \leq S + 1.$$

$$G(s) = \int_{\tau=0}^{s} 10^\tau \ln(10)\eta(\tau)d\tau.$$

So

$$G(s) = const + a_0 s - \frac{a_1}{2\pi}\cos(2\pi s) + \frac{b_1}{2\pi}\sin(2\pi s) + \ldots$$

and

$$\begin{aligned}&\rho(k) \approx G(s_1) - G(s), k \in \mathbb{Z}^*,\\ &\quad s_1 = \log(k+1) - \lfloor \log(k) \rfloor . s = \log(k) - \lfloor \log(k) \rfloor .\end{aligned}$$

In many cases, taking $S = 4$ gives pretty good results.

10 Conclusions

The main results of this paper are as follows.

1. A general form for the DSLD is presented (10).
2. It is shown that in a number of cases this general solution takes the simpler forms (8) and (9).
3. It is proved that instead of the NBL it is natural to look for, in permissible cases, a DSLD of the form (9).
4. The relation between the DSLD and the density of the underlying distribution is presented in the form of an integral relation.

5. The function $P(k,x)$ (18) is introduced, which is convenient for isolating the interval of the leading digits.

6. There has been presented some exact solutions and their relation to the general form of the DSLD.

7. In a number of cases, a Fourier representation is found that can be used to find approximate solutions by cutting short its infinite series. In many practical cases, only the first few terms are sufficient. These approximate analytical formulas provide a high degree of accuracy.

8. It is shown that knowing the results of a numerical experiment for sufficiently large leading numbers, it is possible to recover the solution for their entire range.

9. In many cases, the function $\Omega(s)$ has a quasipolynomial form (with constants α_i and β_{ij}):

$$\Omega(s) = \sum_i 10^{\alpha_i s} \sum_j \beta_{ij} s^j$$

We have found the forms of the DSLD for a set of underlying distributions and presented closed forms for the DSLD for a few of them. We have introduced an integral representation of the DSLD as a function of the underlying distribution.

From these results, we conclude that the logarithmic terms are native terms with sums with other oscillating terms. In many cases, their contribution is essential. Regarding the NBL, it is in many cases a better or worse approximation of more complicated forms of distributions, and can be exact only in a few rare cases, which neglect the oscillating terms.

We found some solutions in the form of an infinite series which with high accuracy can be substituted by the sum of its first few terms. We are confident that our approach can be applied to the analysis of many applied problems. In some cases, it may be preferable to use this approach, but not, for instance, Monte Carlo methods, which are known for their properties of slow convergence.

Here we can formulate a necessary condition for the realization of the NBL

$$\left| \frac{1}{\pi n} \int\limits_{x=0}^{+\infty} x^{Y\imath} f(x) dx \right| \ll \int\limits_{x=0}^{\infty} f(x) dx, Y = \frac{2\pi n}{\ln(10)}, n \in \mathbb{N}. \tag{86}$$

From a practical point of view, the ratio of an (oscillating part)/(log part) should be $o(10^{-3})$.

This approach can be applied for different radices (bases), and then

$$\rho(k,B) = \Omega(\log_B(k+1) - \lfloor \log_B(k) \rfloor) - \Omega\left(\log_B(k) - \lfloor \log_B(k) \rfloor\right),$$
$$\log_B(k) = \frac{\ln(k)}{\ln(B)}, 1 < B, B \text{ positive integer.}$$

APPENDICES

Now we will look in more detail at some of the technical aspects

A Solution of the Functional Equations

Let's temporarily forget about probabilities and consider (87,88) as functional equations with specific symmetry properties. For the PDF of the leading digits we have

$$k \in [10^L, 10^{L+1}-1], L \in \mathbb{Z}^*,$$
$$L = \lfloor \log(k) \rfloor, 0 \le \rho(k, L) \le 1.$$

$$\sum_{k=10^L}^{10^{L+1}-1} \rho(k) = F. \tag{87}$$

$$\rho(k) = \sum_{i=0}^{10^L-1} \rho(10^L k + i, L), L \in \mathbb{Z}^*. \tag{88}$$

A.1 Partial solutions

Using equations (87,88) we can find the partial solutions. Now we will consider different types of solutions of the equations.

- $\rho(k) = g(\lfloor \log(k) \rfloor)$

Consider equation (5) in form

$$\rho(k) = \sum_{i=0}^{9} \rho(10k + i). \tag{89}$$

Then because

$$\left\lfloor \log\left(k + \frac{i}{10}\right) \right\rfloor = \lfloor \log(k) \rfloor, 0 \le i \le 9.$$

$$g(s) = 10g(s+1), s = . \lfloor \log(k) \rfloor.$$

Solution of this equation is

$$g(s) = A10^{-s}.$$

After substitution in (7) with $L = 0$,we have

$$A = \frac{F}{9}.$$

.and

$$g(L) = 10^{-L}\frac{F}{9}, \quad (90)$$
$$\rho(k) = \frac{F}{9}10^{-\lfloor \log(k) \rfloor}.$$

- $\rho(k, L) = f(k)$

Then

$$f(k) = \sum_{i=0}^{10^L-1} f(10^L k + i), L = 0, 1, \dots. \quad (91)$$

Bearing in mind (2), we will consider solutions of the form

$$f(k) = g(k+1) - g(k). \quad (92)$$

Then for $L = 1$

$$g(k+1) - g(k) = g(10(k+1)) - g(10k),$$
$$g(10k) - g(k) = g(10(k+1)) - g(k+1).$$

then

$$g(10k) - g(k) = g(10(k+1)) - g(k+1) = A = const,$$
$$k \in \mathbb{Z}.$$

We get the functional equation

$$g(10k) - g(k) = A. \quad (93)$$

Let us solve the functional equation (93). Introduce the parameter τ and two unknown functions $\psi(\tau)$ and $\varphi(\tau)$

$$k = \varphi(\tau), 10k = \varphi(\tau+1), \quad (94)$$
$$g(k) = \psi(\tau), g(10k) = \psi(\tau+1).$$

$$10\varphi(\tau) = \varphi(\tau+1), \quad (95)$$
$$\psi(\tau+1) - \psi(\tau) = A.$$

The solutions of the difference equations are

$$\varphi(\tau) = 10^{\tau} P_1(\tau), P_1(\tau+1) = P_1(\tau), \forall\tau,$$
$$\psi(\tau) = A\tau + Q_1(\tau), Q_1(\tau+1) = Q_1(\tau), \forall\tau.$$

where $P_1(\tau)$ and $Q_1(\tau)$ are arbitrary functions with period 1. Then we have a parametric solution of (93)

$$
\begin{aligned}
k &= 10^{\tau} P_1(\tau), \\
g(k) &= A\tau + Q_1(\tau).
\end{aligned} \tag{96}
$$

In the partial case when

$$P_1(\tau) = 1, Q_1(\tau) = 0, A = 1.$$

we have the NBL distribution

$$\rho(k) = \log(k+1) - \log(k). \tag{97}$$

If put $P_1(\tau) = const$ we have

$$
\begin{aligned}
\tau &= \log(k) + c, \\
g(k) &= A\log(k) + Ac + Q_1(\log(k) + c),
\end{aligned}
$$

$$\rho(k) = A\left(\log(k+1) - \log(k)\right) + Q_1(\log(k+1) + c) - Q_1(\log(k) + c). \tag{98}$$

From (87)

$$
\begin{aligned}
\sum_{k=10^L}^{10^{L+1}-1} \rho(k) &= \digamma, \\
A + Q_1(L+1+c) - Q_1(L+c) &= \digamma, \\
A &= \digamma.
\end{aligned} \tag{99}
$$

$$\rho(k) = \digamma\left(\log(k+1) - \log(k)\right) + Q_1(\log(k+1) + c) - Q_1(\log(k) + c). \tag{100}$$

If $1 \ll k$, we have

$$\rho(k \to \infty) = \frac{1}{\ln(10)}\frac{1}{k} + \frac{1}{\ln(10)}\frac{1}{k}\left.\frac{dQ_1(s)}{ds}\right|_{s=\log(k)+c} + o\left(\frac{1}{k}\right). \tag{101}$$

If we have some approximation (from experimental data and/or from numerical modeling) $f(s)$ where $\hat{s} \le s \le \hat{s}+1$ and $3 \le \hat{s}$,
then

$$
\begin{aligned}
&G(s) = -s + \ln(10)\int^{s} 10^{t} f(t)dt, \\
&\rho(k) \approx -\log\left(\frac{k+1}{k}\right) + \ln(10)\int_{s}^{s_1} 10^{t} f(t)dt, \\
&s_1 = \log(k+1) + c, s = \log(k) + c, k = 1, 2, \ldots.
\end{aligned}
$$

- $\rho(k, L)$

After some algebra similar to an earlier item, we find

$$\rho(k) = \Omega(\log(k+1) - \lfloor \log(k) \rfloor) - \Omega(\log(k) - \lfloor \log(k) \rfloor). \tag{102}$$

where $\Omega(s)$ is a function of s subject to the restriction

$$0 \leq \rho(k, L) \leq 1.$$

Direct substitution of (102) in (88) give

$$\begin{aligned} &\Omega(\log(k+1)) - \Omega(\log(k)) \\ &= \sum_{i=0}^{10^L-1} \big(\Omega(\log(10^L k + i + 1) - L) - \Omega(\log(10^L k + i) - L)\big) \\ &= \Omega(\log(k+1)) - \Omega(\log(k)), L = 0, 1, \ldots. \end{aligned} \tag{103}$$

and from (87)

$$\begin{aligned} &\sum_{i=10^L}^{10^{L+1}-1} (\Omega(\log(i+1) - L) - \Omega(\log(i) - L)) \\ &= \Omega(1) - \Omega(0) = F, \\ &L = 0, 1, \ldots. \end{aligned} \tag{104}$$

For the distributions of the digits in the second and third positions we have

$$\rho_2(m) = \sum_{i=1}^{9} \rho(10i + m),$$
$$\rho_3(m) = \sum_{i=1}^{9} \sum_{j=0}^{9} \rho(100i + 10j + m), m = 0, 1, \ldots, 9.$$

For a solution of (90) we have

$$\Omega(s) = C 10^s.$$

and for (98)

$$\Omega(s) = C \log(s) + Q_1(s).$$

B From Sum to Integral in (2)

Euler–Maclaurin's formula (**EMf**)[4] ,[5] and [9] gives a representation of sums in form of definite integrals. It is important to stress that the **EMf** is an exact representation of the sum in integral form (at least for differentiable functions, it is not an approximation). From **EMf** we have a representation of a sum in the form of an integral by (see Appendix F)

$$\sum_{j=a}^{b} g(j) = \int\limits_{x=a}^{b} \left(g(x) + \left(x - \lfloor x \rfloor - \frac{1}{2} \right) \frac{dg(x)}{dx} \right) dx + \frac{g(b)}{2} + \frac{g(a)}{2}, a < b. \quad (105)$$

We will apply Euler–Maclaurin summation in this form, which is convenient for our purposes.

In our case we have

$$g(j) = F(10^{j+\log(k+1)}) - F(10^{j+\log(k)})$$
$$F(z) = \int\limits_{-\infty}^{z} f(t) dt.$$

We can, without loss of generality, put infinite limits of integration, since it is always possible to represent the density distribution $\tilde{f}(t)$ as

$$f(t) = \begin{cases} 0, & t_{\max} < t, \\ \tilde{f}(t) & t \in [t_{\min}, t_{\max}] \\ 0 & t < t_{\min}. \end{cases} \quad (106)$$

Let us estimate the sums and integrals in (2).

B.1 Infinite limits

We will consider the case where $j_{\min} = -\infty$ and $j_{\max} = \infty$. From the Euler–Maclaurin formula, for convention sum to integral we have

$$\sum_{j=-\infty}^{\infty} g(j) = \int\limits_{x=-\infty}^{\infty} \left(g(x) + \frac{dg(x)}{dx} \left(x - \lfloor x \rfloor - \frac{1}{2} \right) \right) dx, g(\pm\infty) = 0, \quad (107)$$
$$S = I + R,$$
$$I = \int\limits_{x=-\infty}^{\infty} g(x) dx, R = \int\limits_{x=-\infty}^{\infty} \left(\frac{dg(x)}{dx} \left(x - \lfloor x \rfloor - \frac{1}{2} \right) \right) dx.$$

With

$$S = \sum_{j=-\infty}^{\infty} \left(F(10^j(k+1)) - F(10^j k) \right),$$
$$F(z) = \int\limits_{x=-\infty}^{z} f(x) dx, f(\pm\infty) = 0.$$

and we have

$$I = \int\limits_{x=-\infty}^{\infty} \left(F(10^x(k+1)) - F(10^x k)\right) dx \tag{108}$$
$$= \int\limits_{x=-\infty}^{\infty} \int\limits_{t=\alpha}^{t=\beta} f(t) dt dx,$$
$$\alpha = 10^x k, \beta = 10^x(k+1).$$

After integration by part we have

$$I = -\int\limits_{x=-\infty}^{\infty} x \left(f(\beta)\frac{d\beta}{dx} - f(\alpha)\frac{d\alpha}{dx} \right) dx.$$

and

$$\int\limits_{x=-\infty}^{\infty} x f(\alpha)\frac{d\alpha}{dx} dx = \int\limits_{s=0}^{\infty} \left(\log(s) - \log(k)\right) f(s) ds$$
$$= -\log(k) \int\limits_{s=0}^{\infty} f(s) ds + \int\limits_{\alpha=0}^{\infty} \log(s) f(s) ds.$$

Then the integral part is

$$I = \left(\log(k+1) - \log(k)\right) \int\limits_{s=0}^{\infty} f(s) ds. \tag{109}$$

For the evaluation of the second term in (107), we have

$$R = \int\limits_{x=-\infty}^{\infty} \left(f(\beta)\frac{d\beta}{dx} - f(\alpha)\frac{d\alpha}{dx} \right) \left(x - \lfloor x \rfloor - \frac{1}{2} \right) dx$$

and

$$\int\limits_{x=-\infty}^{\infty} f(\alpha)\frac{d\alpha}{dx} \left(x - \lfloor x \rfloor - \frac{1}{2} \right) dx$$
$$= \int\limits_{s=0}^{\infty} f(s) \left(\log\left(\frac{s}{k}\right) - \left\lfloor \log\left(\frac{s}{k}\right) \right\rfloor - \frac{1}{2} \right) ds,$$

then the residual part is

$$R = -\int_{s=0}^{\infty} f(s)\left(\log(k+1) - \log(k) - \left\lfloor \log\left(\frac{s}{k+1}\right)\right\rfloor + \left\lfloor \log\left(\frac{s}{k}\right)\right\rfloor\right) ds \tag{110}$$

$$= -\left(\log(k+1) - \log(k)\right)\int_{s=0}^{\infty} f(s)ds + \int_{t=0}^{\infty} f(t)\left(\left\lfloor \log\left(\frac{t}{k}\right)\right\rfloor - \left\lfloor \log\left(\frac{t}{k+1}\right)\right\rfloor\right) dt.$$

We obtain the general form of $\rho(k)$ as

$$\rho(k) = \int_{s=0}^{\infty} f(t)\left(\left\lfloor \log\left(\frac{t}{k}\right)\right\rfloor - \left\lfloor \log\left(\frac{t}{k+1}\right)\right\rfloor\right) dt. \tag{111}$$

From this we have an integral representation of $\Omega(s)$ (provided the integral exists)

$$\Omega(s) = -\int_{t=0}^{\infty} f(t)\left\lfloor \log(t) - s\right\rfloor dt + const \tag{112}$$

Another form is

$$\rho(k) = F\left(\log(k+1) - \log(k)\right) + \tag{113}$$

$$\int_{s=0}^{\infty} f(s)\left(\log\left(\frac{s}{k+1}\right) - \left\lfloor \log\left(\frac{s}{k+1}\right)\right\rfloor\right) ds - \int_{s=0}^{\infty} f(s)\left(\log\left(\frac{s}{k}\right) - \left\lfloor \log\left(\frac{s}{k}\right)\right\rfloor\right) ds,$$

$$F = \int_{s=0}^{\infty} f(s)ds. \tag{114}$$

If we introduce the 1-periodic function $Q_1(s)$

$$Q_1(\log(k)) = \int_{s=0}^{\infty} f(t)\left(\log\left(\frac{t}{k}\right) - \left\lfloor \log\left(\frac{t}{k}\right)\right\rfloor\right) dt, \tag{115}$$

$$Q_1(s+1) = Q_1(s).$$

$$Q_1(\log(k)) = \int_{x=-\infty}^{\infty} \ln(10)10^x k f(10^x k)\left(x - \lfloor x\rfloor\right) dx, \tag{116}$$

$$Q_1(s+1) = Q_1(s).$$

and because

$$0 \leq \left(\log\left(\frac{s}{k}\right) - \left\lfloor \log\left(\frac{s}{k}\right)\right\rfloor\right) \leq 1$$

then

$$|Q_1(\log(k))| \leq \digamma.$$

Then we can rewrite (113) as

$$S = \rho(k) = \digamma\left(\log(k+1) - \log(k)\right) + Q_1(\log(k+1)) - Q_1(\log(k)). \quad (117)$$

B.2 Fourier Expansion

Let us consider the 1-periodic function $Q_1(s)$ (115)

$$Q_1(s) = \int\limits_{x=0}^{\infty} f(x)\left(\log(x) - s - \lfloor \log(x) - s \rfloor\right) dx. \quad (118)$$

We will consider the Fourier expansion

$$Q_1(s) = const + \sum_{j=1}^{\infty}\left(A_j \sin(2\pi j s) + B_j \cos(2\pi j s)\right).$$

Because

$$a_n(c) = \int\limits_{s=0}^{1} (c - s - \lfloor c - s \rfloor) \sin(2\pi n s)\, ds \left(\int\limits_{s=0}^{1} \sin(2\pi n s)^2\, ds\right)^{-1},$$
$$b_n(c) = \int\limits_{s=0}^{1} (c - s - \lfloor c - s \rfloor) \cos(2\pi n s)\, ds \left(\int\limits_{s=0}^{1} \cos(2\pi n s)^2\, ds\right)^{-1}.$$

and

$$(\log(x) - s - \lfloor \log(x) - s \rfloor) = a_0 + \sum_{n=1}^{\infty}\left(a_n \sin(2\pi n s) + b_n \cos(2\pi n s)\right)$$
$$a_n = \frac{\cos(2\pi n \log(x))}{n\pi}, b_n = -\frac{\sin(2\pi n \log(x))}{n\pi}, 1 \leq n.$$

we have then

$$\mathcal{F} = \int\limits_{x=0}^{\infty} f(x)dx, \tag{119}$$

$$A_n = \frac{1}{n\pi} \int\limits_{x=0}^{\infty} f(x)\cos(2\pi n \log{(x)})dx, \tag{120}$$

$$B_n = -\frac{1}{n\pi} \int\limits_{x=0}^{\infty} f(x)\sin(2\pi n \log{(x)})dx, 1 \le n.$$

in another form

$$A_n + \imath B_n = \frac{1}{n\pi} \int\limits_{x=0}^{\infty} f(x)\exp(-2\pi n \log{(x)}\,\imath)dx \tag{121}$$

$$= \frac{1}{n\pi} \int\limits_{x=0}^{\infty} x^{-\imath Y} f(x)dx,$$

$$Y = \frac{2\pi n}{\ln(10)}.$$

or in complex polar form we have

$$A_n + \imath B_n = \Lambda(n)\exp(\imath\varphi(n)) = \frac{1}{n\pi} \int\limits_{x=0}^{\infty} x^{-\imath Y} f(x)dx.$$

Then

$$\rho(k) = \mathcal{F}(\log{(k+1)} - \log{(k)}) + \sum_{n=1}^{\infty} \begin{pmatrix} A_n\,(\sin{(ns_1)} - \sin{(ns)}) \\ +B_n(\cos{(ns_1)} - \cos{(ns)}) \end{pmatrix}, \tag{122}$$

$$s_1 = 2\pi \log{(k+1)}\,, s = 2\pi \log{(k)}\,.$$

then

$$\rho(k) = \mathcal{F}(\log{(k+1)} - \log{(k)}) + \sum_{n=1}^{\infty} \Lambda(n)\left(\sin{(ns_1 + \varphi(n))} - \sin{(ns + \varphi(n))}\right). \tag{123}$$

C Ratio of Two Positive Numbers with a UD

Because

$$f_Z(z) = \int \delta\left(z - g(x_1 \ldots x_n)\right) dx_1 \ldots dx_n.$$

where δ is the Dirac delta function, we can find the PDF for different transformations of stochastical variables. Let us assume that $f_X(x)$ is a PDF for $0 \le x$ and $f_Y(y)$ is a PDF for $0 \le y$. We need to find the density of $z = \frac{y}{x}$. If the two random variables are independent, then

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

The distribution function for Z is

$$F(z) = \int^{z} f_Z(s) ds.$$

Let us transform the variables

$$x = \alpha, y = \beta\alpha.$$

Then the Jacobian of the transformation is

$$J = \frac{\partial (x, y)}{\partial (\alpha, \beta)} = \det \begin{pmatrix} \frac{\partial x}{\partial \alpha} & \frac{\partial y}{\partial \alpha} \\ \frac{\partial x}{\partial \beta} & \frac{\partial y}{\partial \beta} \end{pmatrix} = \det \begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix} = \alpha.$$

in case ($x \in (-\infty \ldots + \infty)$ and $y \in (-\infty \ldots + \infty)$)

$$\begin{aligned} F(z) &= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} \alpha f_{XY}(\alpha, z\alpha) d\alpha \right] d\beta \\ &= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} \alpha f_X(\alpha) f_Y(z\alpha) d\alpha \right] d\beta \end{aligned}$$

and then

$$f_Z(z) = \int_{-\infty}^{\infty} \alpha f_{XY}(\alpha, z\alpha) d\alpha$$

C.1 Uniform Distributions

$$f_X(x) = \begin{cases} \frac{1}{b_x - a_x}, & x \in [a_x, b_x] \\ 0, & x \notin [a_x, b_x] \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{b_y - a_y}, & y \in [a_y, b_y] \\ 0, & y \notin [a_y, b_y] \end{cases}$$

$$\frac{a_y}{b_x} \leq z \leq \frac{b_y}{a_x}, 0 < a_x.$$

$$f_Z(z) = \frac{1}{(b_x - a_x)(b_y - a_y)} \int\limits_{\max(a_x, z^{-1}a_y)}^{\min(b_x, z^{-1}b_y)} x dx$$

$$= \frac{\min(b_x, z^{-1}b_y)^2 - \max(a_x, z^{-1}a_y)^2}{2(b_x - a_x)(b_y - a_y)},$$

$$\int\limits_{z=\frac{a_y}{b_x}}^{z=\frac{b_y}{a_x}} f_z(z)dz = 1, 0 < a_x.$$

$$f_z(z) = \begin{cases} \frac{(b_x - a_x)}{2(b_y - a_y)} & 0 < z, \min(b_x, z^{-1}b_y) = b_x, \max(a_x, z^{-1}a_y) = a_x, \\ & \frac{a_y}{a_x} < z < \frac{b_y}{b_x} \\ \frac{b_y^2 z^{-2} - a_x^2}{2(b_x - a_x)(b_y - a_y)} & \min(b_x, z^{-1}b_y) = z^{-1}b_y, \max(a_x, z^{-1}a_y) = a_x, \\ \frac{z^{-2}(b_y - a_y)}{2(b_x - a_x)} & \min(b_x, z^{-1}b_y) = z^{-1}b_y, \max(a_x, z^{-1}a_y) = z^{-1}a_y, \\ 0, & \text{otherwise.} \end{cases}$$

and

$$f_Z(z) = \frac{\min(b_x, z^{-1}b_y)^2}{2b_x b_y}, a_x = 0, a_y = 0,$$

$$f_z(z) = \begin{cases} \frac{1}{2\delta} & 0 < z, z <= \delta, \delta = \frac{b_y}{b_x}, \\ \frac{\delta}{2} z^{-2} & \delta <= z. \end{cases},$$

$$F(z) = \int\limits_0^z f_z(z)dz = \begin{cases} \frac{z}{2\delta} & 0 < z, z <= \delta, \\ 1 - \frac{\delta}{2} z^{-1} & \delta <= z. \end{cases}$$

For example

$$a_x = 1, b_x = 2, a_y = 3, b_y = 7.$$

$$f_Z(z) = \begin{cases} \frac{1}{2} - \frac{9}{8z^2} & \frac{3}{2} < z \leq 3 \\ \frac{3}{8} & 3 < z \leq \frac{7}{2} \\ \frac{49}{8z^2} - \frac{1}{8} & \frac{7}{2} < z < 7 \\ 0, & \text{otherwise.} \end{cases}$$

$$\Omega(s) = \begin{cases} 0, & s \leq 0 \\ \frac{1}{16}, & s = \log(2) \\ \frac{3}{8}, & s = \log(3) \\ \frac{23}{32}, & s = \log(4) \\ \frac{9}{10}, & s = \log(5) \\ \frac{47}{48}, & s = \log(6) \\ 1, & \text{otherwise.} \end{cases}$$

$\rho(k) = \Omega(\log(k+1)) - \Omega(\log(k)), \Omega(1) - \Omega(0) = 1.$

k	1	2	3	4	5	6	7	8	9
$\rho(k)$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{11}{32}$	$\frac{29}{160}$	$\frac{19}{240}$	$\frac{1}{48}$	0	0	0

Table 12: Table Caption 12

D The Product of Two Random Positive Numbers from a UD

Let $f_X(x)$ be a PDF for $0 \leq x$ and $f_Y(y)$ a PDF for $0 \leq y$. We need to find the density of $z = xy$. If the two random variables are independent, then in this case

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

The distribution function for Z is

$$F(z) = \int^{z} f_Z(s) ds.$$

Let us transform the variables

$$x = \alpha, y = \frac{\beta}{\alpha}.$$

Then the Jacobian of the transformation is

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial \alpha} & \frac{\partial y}{\partial \alpha} \\ \frac{\partial x}{\partial \beta} & \frac{\partial y}{\partial \beta} \end{pmatrix} = \det \begin{pmatrix} 1 & -\frac{\beta}{\alpha^2} \\ 0 & \frac{1}{\alpha} \end{pmatrix} = \frac{1}{\alpha}.$$

$$f_z(z) = \int_{-\infty}^{+\infty} \frac{1}{\alpha} f_X(\alpha) f_Y\left(\frac{z}{\alpha}\right) d\alpha$$

In the case of a uniform distribution, we have

$$f_z(z) = \frac{\ln\left(\min\left(b_x, \frac{z}{a_y}\right)\right) - \ln\left(\max\left(a_x, \frac{z}{b_y}\right)\right)}{(b_x - a_x)(b_y - a_y)}.$$

$$a_x a_y \leq z \leq b_x b_y.$$

$$\int_{z=a_x a_y}^{z=b_x b_y} f_z(z) dz = 1.$$

When

$$a_x = 0, a_y = 0,$$
$$f_z(z) = -\frac{\ln\left(\frac{z}{b_x b_y}\right)}{b_x b_y}, 0 \le z \le b_x b_y$$

For example

$$a_x = 1, b_x = 2, a_y = 3, b_y = 7.$$

$$f_z(z) = \begin{cases} \frac{\ln(z)-\ln(3)}{4} & 3 < z \le 6 \\ \frac{\ln(2)}{4} & 6 < z \le 7 \\ \frac{\ln(14)-\ln(z)}{4} & 7 < z \le 14 \\ 0, & \text{otherwise.} \end{cases}$$

$$\int_{z=3}^{z=14} f_z(z)dz = 1.$$

For this distribution, the leading digits are $[1, 3, 4, 5, 6, 7, 8, 9]$, and the probability of the digit 2 is 0.

k	1	2	3	4	5	6	7	8	9
$\rho(k)$	.1587	0	.0337	.1009	.1511	.1733	.1563	.1250	.0970

Table 13: Table Caption 13

D.1 Leading significant digits

$$a_x = 0, a_y = 0, b_x = 1, b_y = 1,$$
$$f_z(z) = -\ln(z), 0 \le z \le 1,$$
$$F(z) = z - z\ln(z).$$

$$\rho(k) = \frac{1}{9}\left(k\ln(k) - (k+1)\ln(k+1) + 1 + \frac{10}{9}\ln(10)\right), 1 \le k \le 9.$$

$$\rho(k, L) = \frac{10^{-L}}{9}\left(k\ln(k) - (k+1)\ln(k+1) + 1 + \frac{(9L+10)}{9}\ln(10)\right),$$
$$10^L \le k \le 10^{L+1} - 1, L = 0, 1, \ldots.$$

$$\sum_{k=10^L}^{10^{L+1}-1} \rho(k, L) = 1, L \in \mathbb{Z}^*.$$

E Product of n Numbers from a UD

We have

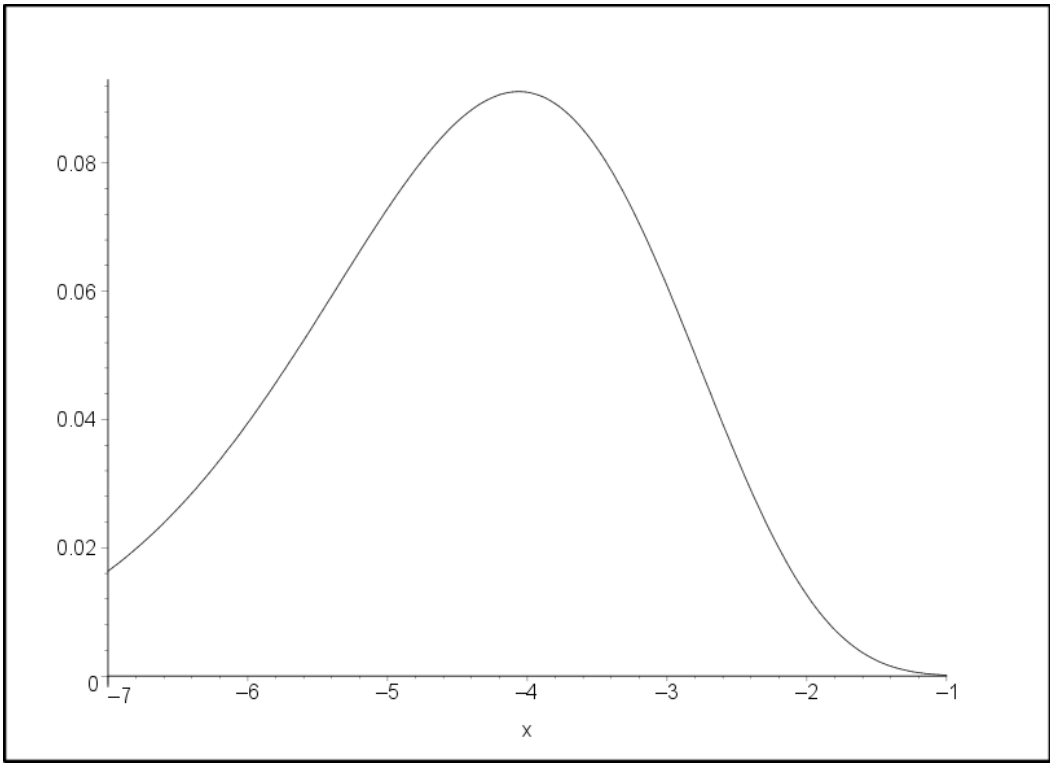
$$f_Z(z,n) = \frac{(-\ln(z))^{in-1}}{(n-1)!},$$
$$F_Z(z,n) = \frac{\Gamma(n,-\ln(z))}{\Gamma(n)} = z\sum_{i=0}^{n}(-1)^i\frac{(\ln(z))^i}{i!}, n \in \mathbb{N}.$$

$$\rho(k,n) = \sum_{m=-\infty}^{-1}(F_Z(10^m(k+1)) - F_Z(10^m k), n)$$
$$= \Omega\left(\log(k+1) - \lfloor\log(k)\rfloor\right) - \Omega\left(\log(k) - \lfloor\log(k)\rfloor\right)$$

From (105) we have

$$a = -\infty, b = -1, \tag{124}$$
$$g(x,k,n) = \frac{\Gamma(n,-\ln(10^x(k+1)))}{\Gamma(n)} - \frac{\Gamma(n,-\ln(10^x k))}{\Gamma(n)},$$
$$I_1(k,n) = \int_{x=-\infty}^{-1} g(x,k,n)dx, I_2 = \int_{x=-\infty}^{-1}\left(x - \lfloor x\rfloor - \frac{1}{2}\right)\frac{dg(x,k,n)}{dx}dx,$$
$$I_3 = \frac{g(-1,k,n)}{2}, I_4 = \frac{g(-\infty,k,n)}{2} = 0.$$

From the graph we have



g(x,1,10) vs x

Figure 11: $g(x, 1, 10)$ vs x

The maximum of $g(x, k, n >> 1)$ is at the point $x = \zeta(n)$, where $\zeta(n >> 1) = -\frac{n}{\ln(10)} + o(n)$. Now we can, without loss of accuracy, change the upper limit to $+\infty$. With the knowledge of (108) and (109),

$$I_1(k, n \to \infty) = \log(k+1) - \log(k), F = 1. \tag{125}$$

We can prove that the integrals as $n \to +\infty$ have the limits

$$I_2(k, +\infty) = 0, I_3(k, +\infty) = 0.$$

This is the NBL Distribution.

F Derivation of Formula (105)

If we have integers a and b ($a < b$), then for an integer j we have, after integration by parts,

$$\int_{j}^{j+1} g(x)dx = \int_{j}^{j+1} g(x)\frac{d\left(x-j-\frac{1}{2}\right)}{dx}dx$$

$$= \frac{g(j+1)}{2} + \frac{g(j)}{2} - \int_{x=j}^{j+1} \left(x-j-\frac{1}{2}\right)\frac{dg(x)}{dx}dx,$$

$$\left(x-j-\frac{1}{2}\right) = x - \lfloor x \rfloor - \frac{1}{2}.$$

Then after summation

$$\sum_{j=a}^{b-1} \int_{j}^{j+1} g(x)dx = \int_{a}^{b} g(x)dx,$$

$$\int_{a}^{b} g(x)dx = \sum_{j=a}^{b} g(j) - \frac{g(b)}{2} - \frac{g(a)}{2} - \int_{x=a}^{b} \left(x - \lfloor x \rfloor - \frac{1}{2}\right)\frac{dg(x)}{dx}dx.$$

Then

$$\sum_{j=a}^{b} g(j) = \int_{a}^{b} \left(g(x) + \left(x - \lfloor x \rfloor - \frac{1}{2}\right)\frac{dg(x)}{dx}\right)dx + \frac{g(b)}{2} + \frac{g(a)}{2}.$$

G Useful Relations

- For any non-negative function $f(x)$ with

  $$0 < F = \int_{0}^{\infty} f(t)dt < const$$

  and positive a and b we have, after integration by parts applied to the left integral,

  $$\int_{-\infty}^{+\infty}\int_{10^x a}^{10^x b} f(t)dtdx = I_b - I_a.$$

$$I_a = -a\ln(10)\int\limits_{-\infty}^{+\infty} x10^x f(10^x a)dx$$

$$I_a = \int\limits_{0}^{+\infty} f(t)(\log(a) - \log(t))dt, x = \log\left(\frac{t}{a}\right),$$

$$I_b = \int\limits_{0}^{+\infty} f(t)(\log(b) - \log(t))dt.$$

$$\begin{aligned} I_b - I_a &= (\log(b) - \log(a))\int\limits_{0}^{+\infty} f(t)dt \\ &= F(\log(b) - \log(a)). \end{aligned}$$

- For any real x and y and integer n we have

$$\lfloor x\rfloor + \lfloor y\rfloor \leq \lfloor x+y\rfloor \leq \lfloor x\rfloor + \lfloor y\rfloor + 1, \tag{126}$$

$$\lfloor x+n\rfloor = \lfloor x\rfloor + n,$$

$$\lceil x\rceil - \lfloor x\rfloor = \begin{cases} 0, & x\in\mathbb{Z}, \\ 1, & x\notin\mathbb{Z}. \end{cases}, \tag{127}$$

$$\lfloor x\rfloor + \lfloor -x\rfloor = \begin{cases} 0, & x\in\mathbb{Z}, \\ -1, & x\notin\mathbb{Z}. \end{cases}. \tag{128}$$

For every real number x and for every positive integer n the following identity holds (Hermite's identity)

$$\lfloor nx\rfloor = \sum_{i=0}^{n-1}\left\lfloor x + \frac{i}{n}\right\rfloor.$$

For a positive integer n, and arbitrary real numbers m and x (x not an integer)$\lfloor$

$$\left\lfloor \frac{\lfloor \frac{x}{m}\rfloor}{n}\right\rfloor = \left\lfloor\frac{x}{mn}\right\rfloor$$

$$\lfloor x\rfloor = x - \frac{1}{2} + \frac{1}{\pi}\sum_{j=1}^{\infty}\frac{\sin(2\pi jx)}{j}. \tag{129}$$

Other presentation

$$\lfloor z\rfloor = z - \frac{1}{2} + \begin{cases} \frac{1}{2}, & z\in\mathbb{Z}, \\ 0. & z = \frac{n}{2}, n\in\mathbb{Z}, \\ \frac{1}{\pi}\arctan\left(\cot(\pi z)\right), & otherwise. \end{cases}$$

or

$$\lfloor z \rfloor = z - \frac{1}{2} + \begin{cases} \frac{1}{2}, & z \in \mathbb{Z}, \\ -\frac{1}{\pi} \arctan\left(\tan\left(\pi\left(z - \frac{1}{2}\right)\right)\right), & z \notin \mathbb{Z}. \end{cases}$$

$$\lceil z \rceil = z + \frac{1}{2} + \begin{cases} -\frac{1}{2}, & z \in \mathbb{Z}, \\ -\frac{1}{\pi} \arctan\left(\tan\left(\pi\left(z - \frac{1}{2}\right)\right)\right), & z \notin \mathbb{Z}. \end{cases}$$

$$\begin{aligned} \cot(x) &= \tan\left(\frac{\pi}{2} - x\right) \\ &= -\tan\left(x - \frac{\pi}{2}\right). \end{aligned}$$

$$\arctan\left(\frac{1}{x}\right) = \operatorname{arccot}(x), x > 0.$$

$$\arctan\left(\frac{1}{x}\right) = \frac{\pi}{2} - \arctan(x), x > 0.$$

$$\arctan\left(\frac{1}{x}\right) = \operatorname{arccot}(x), x < 0.$$

$$\arctan\left(\frac{1}{x}\right) = -\frac{\pi}{2} - \arctan(x), x > 0.$$

$$\arctan(x) = 2\arctan\left(\frac{x}{1+\sqrt{1-x^2}}\right),$$

$$\arctan(x) = \arcsin\left(\frac{x}{\sqrt{1-x^2}}\right),$$

$$\tan(a+b) = \frac{\tan(a)+\tan(b)}{1-\tan(a)\tan(b)}.$$

$$\arctan(-x) = -\arctan(x).$$

$$\frac{d\arctan(x)}{dx} = \frac{1}{1+x^2}.$$

$$\arctan(z) = \frac{\imath}{2}\left(\ln(1-\imath z) - \ln(1+\imath z)\right).$$

$$\arctan(a) - \arctan(b) = \arctan\left(\frac{a-b}{1+ab}\right).$$

$$\tan(a+b) = \arctan\left(\frac{\tan(a)+\tan(b)}{1-\tan(a)\tan(b)}\right).$$

From RHS of (129) we have

$$x - \frac{1}{2} + \frac{1}{\pi}\sum_{j=1}^{\infty}\frac{\sin(2\pi jx)}{j} \tag{130}$$

$$= x - \frac{1}{2} + \sum_{j=1}^{\infty}\frac{1}{2\pi j\imath}\left(\exp(2\pi jx\imath) - \exp(-2\pi jx\imath)\right). \tag{131}$$

after summation we have

$$\sum_{j=1}^{\infty}\frac{\sin(2\pi jx)}{j} = \frac{1}{2\pi\imath}\left(-\ln\left(1-\exp(-2\pi x\imath)\right) + \ln\left(1-\exp(2\pi x\imath)\right)\right)$$
$$= \frac{1}{\pi}\arctan(\cot(\pi x)), x \notin \mathbb{Z}.$$

G.1 Code

Here is R code

```
V<-function(s,n){if(s==floor(s)) return(s);
j<-c(1:n); a<-s-1/2+sum(sin(2*pi*j*s)/(pi*j));
return(a)}

VV<-function(k,x,n){s<-log10(x/k);
s1<-log10(x/(k+1));
V(s,n)-V(s1,n)}

VVV<-function(k,Sample,n){N<-length(Sample);
for(j in 1:N){a[j]<-VV(k,Sample[j],n)};
return(sum(a)/N)}
```

Where `Sample` is list of positive numbers, `n` is positive integer.

Acknowledgement 1 *I am grateful to Prof. T. Hill for pointing out to me the Benford Online Bibliography. I thank Dr. F. Benford, Dr. S. Miller and Alex E. Kossovsky for useful discussions and comments.*

References

[1] S.Newcombs *Note on the frequencey of use of the different digits in natural numbers* 1881: Amer. J. Math. 4 39-40.

[2] F.Benford *The law of anomalous numbers.* 1938: Proceedings of the American Philosophical Society. 78 551-572.

[3] R.S. Pinkham *On the Distribution of First Significant Digits.* 1961: Ann. Math. Statist. Volume 32, Number 4, 1223-1230.

[4] D.Knuth, *The Art of Computer Programming* 1965: 2 219-229. Addison-Wesley, Reading, MA.

[5] M.Abramowitz, I.A.Stegun (Eds.)*Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*,9th printing. 1972: New York: Dover, pp. 16 and 806.

[6] P.Diaconis *The distribution of leading digist and uniform distribution mod 1.* 1977: Ann. Probab. 5 72-81.

[7] R.Raimi *The first digit phenomenon again.* 1985: Proceedings of the American Philosophical Society 129 211-219.

[8] M.J.Nigrini *The Detection of Income Tax Evasion Through an Analysis of Digital Distributions* 1993. Ph.D. Dissertation, Univ. of Cincinnatti.

[9] R.L. Graham, D.E. Knuth,O.Patashnik *Concrete Mathematics: A Foundation for Computer Science. 2nd Edition* 1994: Addison-Wesley Professional, pp.672.

[10] T.Hill, *The significant-digit phenomenon* 1995: Amer. Math. Monthly 102 322-327.

[11] M.J.Nigrini *A Taxpayer Compliance Application of Benford's Law.* 1996: Journal of the American Tax Association, Spring 1996 pp 72-91.

[12] V.I.Arnold *Statistics of first digits of degries of 2 and redivision of world.* 1998: Kvant, 1.(in Russian)

[13] V.I.Arnold *The antiscientifical revolution and mathematics* Talk at the meeting of the Pontifical Academy at Vatican, 26 October 1998, Changing concepts of nature at the turn of the millennium.

[14] C.Walck, *Handbook on Statistical Distributions for experimentalists* 2007. Particle Physics Group, Fysikum University of Stockholm.. http://www.stat.rice.edu/ dobelman/textfiles/DistributionsHandbook.pdf

[15] A.K.Formann, Richard James ed. *"The Newcomb-Benford Law in Its Relation to Some Common Distributions".* 2010:PLoS ONE. 5 (5): e10541. Bibcode:2010PLoSO. 510541F. doi:10.1371/journal.pone.0010541. PMC 2866333free to read. PMID 2047987

[16] HA.Berger,T.P.Hill, *Benford Online Bibliography* 2009: accessed May 14, 2010, at h http://www.benfordonline.net.

[17] M.Nigrini *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection .* 2012:Princeton University Press , ISBN: 978-1118152850 330 pp.

[18] Arno Berger , Theodore P. Hill *An Introduction to Benford's Law.* 2015: Princeton University Press SBN: 9780691163062 256 pp.

[19] Steven J. Miller ed. *Benford's Law: Theory and Applications.* 2015:Princeton University Press , ISBN: 978-0-691-14761-1: 464 pp.