Dirty Pages of Logarithm Tables, Lifetime of the Universe, and (Subjective) Probabilities on Finite and Infinite Intervals

Hung T. Nguyen[1], Vladik Kreinovich[2], and Luc Longpré[2]

[1]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
email hunguyen@nmsu.edu
[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
emails {vladik,longpre}@cs.utep.edu

Abstract

To design data processing algorithms with the smallest average processing time, we need to know what this ''average" stands for. At first glance, it may seem that real-life data are really "chaotic", and no probabilities are possible at all: today, we may apply our software package to elementary particles, tomorrow – to distances between the stars, etc. However, contrary to this intuitive feeling, there are stable probabilities in real-life data. This fact was first discovered in 1881 by Simon Newcomb who noticed that the first pages of logarithm tables (that contain numbers starting with 1) are more used than the last ones (that contain numbers starting with 9). To check why, he took all physical constants from a reference book, and counted how many of them start with 1. An intuitive expectation is that all 9 digits should be equally probable. In reality, instead of 11%, about 30% of these constants turned out to be starting with 1. In general, the fraction of constants that start with a digit d can be described as $\ln(d+1) - \ln(d)$. We describe a new interval computations-related explanation for this empirical fact, and we explain its relationship with lifetime of the Universe and with the general problem of determining subjective probabilities on finite and infinite intervals.

1 The Need for Subjective Probability on Finite Intervals

In many engineering problems, we want a physical characteristic y to lie within given range $\mathbf{Y}$; e.g., a stress y of a mechanical structure should not exceed a given value y_0; a temperature y within a chemical reactor should not exceed a critical value y_0 after which the walls become damaged, etc. For most such problems, we know the dependence $y = f(x_1, \ldots, x_n)$ of this characteristic y on the design parameters $x_1, \ldots, x_n$, and we know the intervals $\mathbf{x}_i$ of possible values of these parameters that correspond to a given design. We can then use interval computations to find the corresponding range of y. If this range is completely within the desired range $\mathbf{Y}$, great.

But what if no such design is possible? In this case, from the purely mathematical viewpoint, none of the proposed designs is completely satisfying, so all of them are equally bad. Intuitively, however, some designs seems to be more "probable" to be good for the actual (unknown) values of the parameters x_i. For example, suppose that we have a single parameter x whose interval of possible values is $[0, 1]$, then, intuitively, a design for which $y = f(x) \in \mathbf{Y}$ for all values $x \in [0.001, 1]$ is more probable to work well than a design for which $y \in \mathbf{Y}$ only for the values $x \in [0, 0.5]$.

How can we describe this subjective notion of probability?

2 Subjective Probability on Finite Intervals: Motivations, Definition, and the Resulting Description (1D Case)

Let's first consider a *1-dimensional case*, i.e., the case when we are describing the value of only one physical quantity x. Our goal is to describe, for the case when we know that the actual value x is within an interval $[a, b]$ (and no other information about x is available), the corresponding (subjective) probability of different values within this interval. For each subset $A \subseteq [a, b]$, the corresponding subjective probability will be denoted by $p_{[a,b]}(A)$.

This value may not be defined for some complex sets A, but we want it to be well-defined at least for every subinterval $[c, d] \subseteq [a, b]$. In other words, we require that once we know that $x \in [a, b]$, and $[c, d] \subseteq [a, b]$ is a subinterval of the interval $[a, b]$, then there is a subjective probability $p_{[a,b]}([c, d])$ that $p \in [c, d]$. Of course, once we know that $x \in [a, b]$, the probability that $x \in [a, b]$ should be 1, i.e., $p_{[a,b]}([a, b]) = 1$.

Since the probability to get values outside $[a, b]$ is 0, in principle, we can define $p_{[a,b]}([c, d])$ for all intervals $[c, d]$, not only for subintervals of $[a, b]$: namely, we can define this probability as the probability for x to be within the intersection $[a, b] \cap [c, d]$.

What are the natural requirements on such probability measures?

The first requirement is *consistency* between different measures. Suppose that initially, our only knowledge about the physical quantity x is that its value belongs to an interval $[a, b]$. Then, we made an additional measurement, and as a result of that measurement, get a smaller interval $[c, d]$ for the same quantity. The initial subjective probability that $x \in [c, d]$ was $p_{[a,b]}([c, d])$. What we did by adding the new knowledge is we deleted the values from the semi-open intervals $[a, c)$ and $(d, b]$ from the list of possible values, so the new probabilities of these values are now 0. We did not, however, provide any new information about the probability of the values inside $[c, d]$. Hence, it is natural to describe the new probabilities $p_{[c,d]}(A)$ as conditional probabilities under the condition $x \in [c, d]$, i.e., to require that $p_{[c,d]}(A) = p(A \,|\, x \in [c, d])$. The conditional probability $P(A|B)$ is defined as $P(A \cap B)/P(B)$. Therefore, for every set A, we have the following requirement:

$$p_{[c,d]}(A) = \frac{p_{[a,b]}(A \cap [c, d])}{p_{[a,b]}([c, d])}.$$

The second natural requirement is *shift-invariance.* The values a, b, etc, are usually obtained by measurements. If we change the starting point of the measurement (time and temperature are good examples where such a change is possible), then all the measured values are shifted ($x \to x + c$ for a fixed c). This is a formal change that does not affect our knowledge, like a change from Kelvin to centigrade in measuring temperature. Therefore, it is natural to assume that the subjective probabilities do not change under this change.

For example, suppose that we know that (in centigrade) the temperature is from the interval $[0, 50]$, and we are interested in the subjective probability that the temperature is actually from the interval [0,20]. This probability is $p_{[0,50]}([0, 20])$. In Kelvin, this same question has a different numerical meaning. Here, the initial information is that $T \in [273, 323]$, and we are interested in the probability that it this temperature is actually in the interval $[273, 293]$. So, the probability of the same event can be described as $p_{[273,323]}([273, 293])$. The two expressions for probability of the same event must coincide, i.e., $p_{[0,50]}([0, 20]) = p_{[273,323]}([273, 293])$.

In general, we must have the shift-invariance condition $p_{[a,b]}(X) = p_{[a+c,b+c]}(X + c)$.

The third requirement is *unit-invariance.* If we change the unit in which we measure the physical quantity (i.e., go from inches to centimeters), then, the numerical values of this quantity change as $x \to \lambda \cdot x$ for some $\lambda > 0$ (= the ratio of the old and the new units). The probabilities must not change under this change either. So, we arrive at the formula $p_{[a,b]}(X) = p_{[\lambda \cdot a, \lambda \cdot b]}(\lambda \cdot X)$.

As a result, we arrive at the following definition:

Definition 1. *By a 1D subjective probability, we mean a function p that to every interval $[a, b]$, puts into correspondence a probability measure $p_{[a,b]}$ with the following properties:*

- *For each interval $[a, b]$, the value $p_{[a,b]}([c, d])$ is defined for all intervals $[c, d]$,*
- *For each interval $[a, b]$, the probability measure $p_{[a,b]}$ is localized on the interval $[a, b]$ (i.e., $p_{[a,b]}([a, b]) = 1$).*
- *Measures $p_{[a,b]}$ that correspond to different intervals $[a, b]$ are* consistent, *i.e.,*

  $$p_{[c,d]}(X) = \frac{p_{[a,b]}(X \cap [c, d])}{p_{[a,b]}([c, d])}$$

  for arbitrary $a, b, c, d \in R$, and for an arbitrary $p_{[a,b]}-$measurable set $X \subseteq R$.
- *The measures $p_{[a,b]}$ are* shift-invariant, *i.e.,*

  $$p_{[a,b]}(X) = p_{[a+c,b+c]}(X + c)$$

  for an arbitrary set X.
- *The measures $p_{[a,b]}$ are* unit-invariant, *i.e.,*

  $$p_{[a,b]}(X) = p_{[\lambda \cdot a, \lambda \cdot b]}(\lambda \cdot X)$$

  for an arbitrary set X.

It turns out that the above requirements uniquely describe subjective probability:

Proposition 1. *If p is a subjective 1-dimensional probability, then*

$$p_{[a,b]}([c, d]) = \frac{|[a, b] \cap [c, d]|}{|[a, b]|},$$

where by $|X|$, we denoted the length of the interval X.

(For readers' convenience, all the proofs are placed in the special Proofs section.)

Comment 1. In Proposition 1, we started with the situation in which we know nothing about the probability of different values $x \in [a, b]$, and we used natural symmetry requirements to uniquely determine these probabilities. The fact that symmetries can help in case of uncertainty is no accident; in our previous interval-related papers, we used symmetry:

- in [16], to find the optimal selection of a side to bisect;
- in [21], to select an optimal formula for the so-called "ε-inflation";
- in [19], to optimally select a sub-box, and
- in [23], for several other computational problems.

Comment 2. We get the same uniform distribution as in Proposition 1 if we use a Maximum Entropy approach (see, e.g., [14, 18, 20]), i.e., select, among all possible probability distribution on the interval $[a, b]$, a distribution with the largest possible entropy $-\int \rho(x) \cdot \log(\rho(x))\,\mathrm{d}x$. This coincidence is not surprising, because the maximum entropy criterion is clearly shift- and unit-invariant.

Comment 3. The use of uniform distributions is also in line with the recommendations of several metrological (= measurement-related) organizations that suggest to use uniform distribution if the only information we have is that the measured value x belongs to an interval $[a, b]$ [4, 5, 33]; see also [3].

3 Applications: 1D Case

The natural uniform distribution has been used to describe subjective probability of different subintervals in numerous areas including:

- earthquake engineering [7], where it is used to gauge the probability with which different design are earthquake-proof;
- technical diagnostics and manufacturing [15, 24, 25], where it is used to describe the probability that the value x of the physical parameter about which we only know that $x \in [a, b]$ actually exceeds the critical value x_0 (when $a < x_0 < b$);
- material science [22], where it is used to select a material that has the largest probability of having thermophysical properties within a desired range;
- metrology [22], where it is used to select a sensor that has the largest probability of covering the desired range of values.

An interesting use of uniform distribution in problems like estimating the lifetime of the Universe comes from R. Gott [10]. Gott's main idea is as follows. Suppose that we are witnessing some process that started at a moment t_s (not necessarily known) and that will end at the moment t_e (also not necessarily known). In accordance with the above result, the current observation time t is uniformly distributed on the interval $[t_s, t_e]$. Therefore, the probability that t happens to be in the first 5% of this interval (i.e., in the interval

$[t_s, t_s + 0.05 \cdot (t_e - t_s)]$), is equal to 5% (5% is just an example, any other small value will do). So, with 95% probability, the current moment of time t is later than $t_s + 0.05 \cdot (t_e - t_s)$.

Suppose now that we know t_s and t, but we do not know t_e. We have already argued that with a 95% probability, $t \geq t_s + 0.05 \cdot (t_e - t_s)$. This inequality leads to $t - t_s \geq 0.05 \cdot (t_e - t_s)$ and $t_e - t_s \leq 20 \cdot (t - t_s)$. In other words, with a 95% probability, the total lifetime $t_e - t_s$ of a process does not exceed 20 times its current age. Examples:

- for the humanity (current age $\approx 200,000$ years), Gott concludes that with a 95% probability, its lifetime will not exceed $20 \cdot 200,000 = 4$ million years;
- for the Universe (current age ≈ 20 billion years), with a 95% probability, its lifetime will not exceed $20 \cdot 20 = 400$ billion years;
- computer era (started in 1994, ≈ 50 years old) will probably last for ≤ 1000 more years, etc.

4 Subjective Probability on Multi-D Finite Intervals: Motivations, Definition, and the Resulting Description

Let's now consider a *multi-dimensional case*, i.e., the case when we are describing the values of several physical quantities $x_1, \ldots, x_n$. In this case, interval information can be described by a *box* $B = [a_1, b_1] \times \ldots \times [a_n, b_n]$.

On each box B, we want to define a probability measure p_B.

Definition 2. *By an n-dimensional subjective probability, we mean a function p that to every box $B = [a_1, b_1] \times \ldots \times [a_n, b_n]$, puts into correspondence a probability measure p_B on R^n that satisfies the following properties:*

- *For each box B, the value $p_B(C)$ is defined for every box C.*
- *For every box B, the measure p_B is localized on the box B (i.e., $p_B(B) = 1$).*
- *The measures p_B that correspond to different sets B are consistent, i.e.,*

$$p_C(X) = \frac{p_B(X \cap C)}{p_B(C)}$$

for arbitrary boxes B, C, and for an arbitrary set X.

- *The measures are shift-invariant, i.e.,*

$$p_B(X) = p_{B+\vec{c}}(X + \vec{c})$$

for an arbitrary vector $\vec{c} \in R^n$.

- *The measures are unit-invariant, i.e.,*

$$p_B(X) = p_{\vec{\lambda}\cdot B}(\vec{\lambda} \cdot X)$$

for an arbitrary vector $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$ with $\lambda_i > 0$. Here, $\vec{\lambda} \cdot X$ means componentwise multiplication of vectors, i.e., $\vec{\lambda} \cdot X \stackrel{\text{def}}{=} \{\vec{\lambda} \cdot \vec{a} \,|\, \vec{a} \in X\}$, and $\vec{\lambda} \cdot \vec{a} \stackrel{\text{def}}{=} (\lambda_1 \cdot a_1, \ldots, \lambda_n \cdot a_n)$.

Motivations for these requirements are similar to the motivations for the 1-dimensional case: shift means changing the starting points of all n quantities, and unit-invariance means changing n measuring units.

Proposition 2. *If p is a subjective n-dimensional probability, then*

$$p_B(C) = \frac{|B \cap C|}{|B|},$$

where by $|X|$, we denoted the n-dimensional measure of the box X (i.e., area for $n = 2$, volume for $n = 3$, etc).

5 Applications: Multi-D Case

1D subjective probabilities are used to compare the value known with interval uncertainty with a threshold.

In many real-life situations, we need to compare interval values with each other. For example, control rule bases often include rules like "if the temperature a is higher than the temperature b, then open valve 1, else open valve 2." In practice, after measurements, we only have intervals $\mathbf{a}$ and $\mathbf{b}$ of possible values of a and b. If the corresponding two intervals intersect, then none of the temperatures is guaranteed to be higher than another.

On a box $\mathbf{a} \times \mathbf{b}$, we have a naturally defined (subjective) probability. A natural idea is therefore to choose an interval for which the probability that $a \geq b$ is greater than the probability that $b \geq a$, i.e., the probability

$$p_{\mathbf{a}\times\mathbf{b}}(\{(a, b) \,|\, a \geq b\})$$

that $a \geq b$ is greater than 1/2.

Another possibility is to take into consideration that the inequality $a \geq b$ is equivalent to $a - b \geq 0$. Since we know that $a \in \mathbf{a}$ and that $b \in \mathbf{b}$, we can conclude that the difference $x \stackrel{\text{def}}{=} a - b$ belongs to the interval $\mathbf{a} - \mathbf{b}$. So, as the desired probability, we can take the conditional probability that the number x is non-negative under the condition that $x \in \mathbf{a} - \mathbf{b}$, i.e., the conditional probability $p_{\mathbf{a}-\mathbf{b}}[0, \infty)$.

It turns out that both ways lead to the same selection:

Proposition 3. *For every two intervals $\mathbf{a} = [\underline{a}, \overline{a}]$ and $\mathbf{b} = [\underline{b}, \overline{b}]$, the following three conditions are equivalent to each other:*

i) $p_{\mathbf{a}-\mathbf{b}}[0, \infty) \geq 1/2$ *for 1-dimensional subjective probability p.*

ii) $p_{\mathbf{a}\times\mathbf{b}}(\{(a, b) \mid a \geq b\}) \geq 1/2$ *for 2-dimensional subjective probability p.*

iii) $\dfrac{\underline{a} + \overline{a}}{2} \geq \dfrac{\underline{b} + \overline{b}}{2}$.

This criterion is actually used in a expert system shell FEST described in [32]. According to this criterion, out of several values known with interval uncertainty, we select a one for which the midpoint is larger (if we are looking for a maximum).

The above idea means that even if we have a 50.1% probability that a is better than b, we reject b and choose a. In many cases, we do not want to make a rejection decision on such weak a basis. So, we may choose a value $p_0 > 1/2$, and reject an alternative c only if there exists another alternative a with $P(a > b) \geq p_0$.

To be able to make these choices, we must be able to compute the corresponding probabilities. The formulas are provided by the following proposition:

Proposition 4.

i) *For 1-dimensional subjective probability p,*

$$p_{\mathbf{a}-\mathbf{b}}[0, \infty) = \frac{\overline{a} - \underline{b}}{\overline{a} - \underline{b} - \underline{a} + \overline{b}}$$

if $\overline{a} \geq \underline{b}$ and 0 else.

ii) *For 2-dimensional subjective probability p,*

$$p_{\mathbf{a}\times\mathbf{b}}(\{(a, b) \mid a \geq b\}) = \frac{I_1 + I_2 + I_3}{(\overline{a} - \underline{a})(\overline{b} - \underline{b})},$$

where $I_1 \stackrel{\text{def}}{=} (1/2)\max(0, \min(\overline{a}, \overline{b}) - \max(\underline{a}, \underline{b}))^2$, $I_2 \stackrel{\text{def}}{=} (\overline{b} - \underline{b}) \cdot \max(0, \overline{a} - \overline{b})$, and $I_3 \stackrel{\text{def}}{=} (\overline{a} - \underline{a}) \cdot \max(0, \underline{a} - \underline{b})$.

Comment 1. Part ii) was first proved in [32].

Comment 2. It is worth mentioning that for $p_0 \neq 1/2$, these formulas are different. Therefore, unless we take $p_0 = 1/2$, the two ways to define probability lead to different sets of solutions. For example, for $\mathbf{a} = [1, 3]$ and $\mathbf{b} = [0, 2]$, the first formula leads to 0.75, and the second one to 0.875.

Comment 3. For the first formula, we can get an explicit criterion for choosing the best alternative a:

Definition 3. *Let $\mathcal{A}$ be a family of intervals. Assume that a real number $p_0 \in [1/2, 1]$ is fixed. We say that an element $\mathbf{a} \in \mathcal{A}$ is the possibly best interval with probability $\geq p_0$ if for every $\mathbf{b} \in \mathcal{A}$ ($\mathbf{b} \neq \mathbf{a}$), the probability $p_{\mathbf{a}-\mathbf{b}}[0, \infty)$ is $\geq p_0$.*

Proposition 5. *For every family $\mathcal{A}$ of intervals, the interval $\mathbf{a} \in \mathcal{A}$ is the possibly largest interval with probability $\geq p_0$ if and only if*

$$p_0 \cdot \underline{a} + (1 - p_0) \cdot \overline{a} \geq \sup_{\mathbf{b} \in \mathcal{A}, \mathbf{b} \neq \mathbf{a}} ((1 - p_0) \cdot \underline{b} + p_0 \cdot \overline{b}).$$

In some cases, the possible values of the objective function do not form an interval; for example, if we have finitely many different possibilities each of which leads to an interval of possible values, then the set of all possible values of $f(a)$ is a *union* of finitely many intervals.

Another case when such a union appears is the case of expert systems [29, 32], when an expert may say that the value of a quantity a belongs to a certain interval $\mathbf{a}$, and that it *does not* belong to another interval $\mathbf{b} \subset \mathbf{a}$. In this case, the resulting knowledge is that a belongs to the set $\mathbf{a} - \mathbf{b} = [\underline{a}, \overline{a}] - [\underline{b}, \overline{b}]$, which is the union of two intervals $[\underline{a}, \underline{b}] \cup [\overline{b}, \overline{a}]$. If we have several negative statements, the resulting set of possible values may be a union of more than two non-intersecting intervals.

In these cases, to compare the choices of a and b, we can compare these sets of possible outcomes. In [32], a probabilistic approach is generalized to this case. Namely, if we know the set A of possible values of a and the set B of possible values of b, then we can define the probability of $a \geq b$ as follows:

- we choose the intervals $\mathbf{a}$ and $\mathbf{b}$ that contain A and B – e.g., as interval hulls of the sets A and B;
- we define the desired probability as the conditional probability that $a \geq b$ on the condition that $a \in A$ and $b \in B$, i.e., as

$$p(A \geq B) \stackrel{\text{def}}{=} \frac{p_{\mathbf{a}\times\mathbf{b}}(\{(a, b) \mid a \in A \,\&\, b \in B \,\&\, a \geq b\})}{p_{\mathbf{a}\times\mathbf{b}}(A \times B)},$$

  where p is a subjective 2-dimensional probability.

This definition uses the intervals $\mathbf{a}$ and $\mathbf{b}$, but the result turn out to be independent on them:

Proposition 6. *For every two intervals $\mathbf{a} \supseteq A$ and $\mathbf{b} \supseteq B$, and for measurable sets A and B, the above formula leads to the same value*

$$P(A \geq B) = \frac{\mu_2(\{(a,b) \in A \times B \,|\, a \geq b\})}{\mu_1(A) \cdot \mu_1(B)},$$

where μ_k denotes a $k-$dimensional Lebesgue measure: length for $k = 1$, and area for $k = 2$.

Comments 1. The proof follows directly from the formulas for subjective 2-dimensional probability.

Comment 2. For the case when both A and B are finite unions of intervals, an explicit formula for $P(A \geq B)$ is given in [32]. Namely, if we combine intersecting intervals into a larger interval that they constitute, we can represent each of the sets A and B as a finite union of non-intersecting intervals. Then, the following formula applies:

Proposition 7. [32] *If*

$$A = \bigcup_{i=1}^{n} \mathbf{a}_i \quad \text{and} \quad B = \bigcup_{j=1}^{m} \mathbf{c}_j$$

(where $\mathbf{a}_i \cap \mathbf{a}_j = \mathbf{b}_i \cap \mathbf{b}_j = \emptyset$ for $i \neq j$), then

$$P(A \geq B) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{L(\mathbf{a}_i, \mathbf{b}_j)}{\mu_1(A) \cdot \mu_1(B)},$$

where for arbitrary two intervals $\mathbf{a}$ and $\mathbf{b}$,

$$L(\mathbf{a}, \mathbf{b}) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \max(0, \min(\overline{a}, \overline{b}) - \max(\underline{a}, \underline{b}))^2 +$$

$$(\overline{b} - \underline{b}) \cdot \max(0, \overline{a} - \overline{b}) + (\overline{a} - \underline{a}) \cdot \max(0, \underline{a} - \underline{b}).$$

The proof is pretty much straightforward (see proof of Proposition 4). For details, one can see [32].

6 Problems with the Above Formulas for Subjective Probability

It is not immediately clear how to generalize this approach to the case when instead of a finite interval, we have an *infinite* (or at least very large) interval, i.e., a semi-line. Let us give an example. Suppose that the condition that we want to satisfy is $a \le a_0$, and that we know that $a \ge 1$. What is the "subjective probability" that the condition $a \le a_0$ is satisfied? I.e, if we use the above notation, what is the probability $p_{[1,\infty)}([1, a_0])$?

An infinite interval $[1,\infty)$ is a limit case of a finite interval $[1, N]$ when $N \to \infty$. So, a natural idea is to apply the above-described approach to compute the probability $p(N) = p_{[1,N]}([1, a_0])$ and then tend to the limit $N \to \infty$. Unfortunately, this idea does not work: due to Proposition 1,
$p(N) = (a_0 - 1)/(N - c)$, and so, in the limit, we get probability 0.

This mathematical argument can be easily reformulated in commonsense terms. Let us assume that we live in an infinite Universe that starts at time 0 and goes on and on. If the Universe is 1 billion years old, then according to Gott's argument, with probability 95%, we are not in its first 50 million year. If the Universe is 100 billion years old, then with the same probability, we cannot be in its first 5 billion years. As we increase the lifetime, these first 5% spread to the entire Universe. We therefore arrive at a counter-intuitive conclusion that with a probability 95%, we cannot be in any time of the Universe. The same conclusion can be made if instead of 95%, we take 99.9%, etc.

This problem with the lifetime of the Universe may look somewhat theoretical: after all, according to modern physics, the Universe is finite. However, there is another example when the above approach does not work well: the problem of dirty pages of logarithm tables. In 1881, Simon Newcomb, a well-known astronomer, noticed that the first pages of logarithm tables (that contain numbers starting with 1) are more used than the last ones (that contain numbers starting with 9) (for a detailed description and references, see [13]). To check why, he took all physical constants from a reference book, and counted how many of them start with 1. If real numbers representing physical constants were distributed uniformly, we would expect all 9 possible first digits appear with the same probability of $\approx 11\%$. In reality, instead of 11%, about 30% of these constants turned out to be starting with 1. In general, the fraction of constants that start with a digit d can be described as $\ln(d+1) - \ln(d)$.

This empirical fact was later rediscovered by F. Benford [2] and is therefore known as *Benford's law.*

A similar law describes not only physical constants, it also describes different types of data ranging from stock exchange to census data to accounting-related numbers. Benford's law is not simply a curious empirical phenomenon, it has been successfully used to, e.g., uncover accounting fraud: actual numbers satisfy this law, while the cooked up data usually follow the uniform distribution. It is therefore important to figure out why this law is so frequent in real life.

This problem is more difficult that one might think because not only the corresponding distribution is different from the seemingly natural uniform distribution, it is difficult to figure out what distribution we have at all. Several authors (see, e.g., [27, 9, 28]) deduced this formula from the requirement similar to our unit-invariance (which they call *scale-invariance*). Crudely speaking, they deduce the formula $p([\underline{a},\overline{a}]) = \text{const} \cdot (\ln(\overline{a}) - \ln(\underline{a}))$. We say "crudely speaking" because $\ln(x) \to \infty$ as $x \to \infty$, so the above formula cannot describe an actual probability distribution; in reality, the authors use some tricks:

- In [27], only the invariance of the digit distribution is required.
- In [9], p is defined as a *limit* of probability measures, and invariance is formulated for this limit – which is not a probability measure.
- In [28], p is defined as a finitely additive measure that is not $\sigma-$additive.
- In [6, 11], the logarithmic distribution is deduced from the following fact: the values of the physical constants are usually obtained by processing data, i.e., by applying several (usually, many) arithmetic operations to the initial data. It turned out that if we start with some random numbers, and apply many (n) arithmetic operations, then as $n \to \infty$, the distribution of the first digit of the result approaches the logarithmic distribution. Therefore, the logarithmic distribution is a good approximation for large n.
- Probably the most mathematically satisfying derivation comes from considering a collection of different probability distributions instead of a single one [12, 13].

In this paper, we describe a new interval computations-related explanation for Benford's law, and we show how this law is related to the general problem of determining subjective probabilities on finite and infinite intervals.

7 Subjective Probability on Infinite Intervals

It turns out that a natural way to avoid the above problems in the infinite case is not to require some of the conditions that we had for finite case. Let's do it for our problem. Namely, we will skip shift-invariance:

Definition 4. *By a 1-dimensional subjective probability on infinite intervals, we mean a function p that to every interval $[a, \infty), a > 0$, puts into correspondence a probability measure $p_{[a,\infty)}$ with the following properties:*

- *For each interval $[a, \infty)$, the value $p_{[a,\infty)}([c, d])$ is defined for all (finite and infinite) intervals $[c, d]$.*
- *For each interval $[a, \infty)$, the probability measure $p_{[a,\infty)}$ is localized on the interval $[a, \infty)$ (i.e., $p_{[a,\infty)}([a, \infty)) = 1$).*
- *Measures $p_{[a,\infty)}$ that correspond to different intervals $[a, \infty)$ are consistent, i.e.,*
$$p_{[c,\infty)}(X) = \frac{p_{[a,\infty)}(X \cap [c, \infty))}{p_{[a,\infty)}([c, \infty))}$$
for arbitrary $a, b \in R$, and for an arbitrary $p_{[a,\infty)}-$measurable set $X \subseteq R$.

- *The measures $p_{[a,\infty)}$ are unit-invariant, i.e.,*

$$p_{[a,\infty)}(X) = p_{[\lambda\cdot a,\infty)}(\lambda \cdot X)$$

for an arbitrary set X.

Proposition 8. *For every subjective 1-dimensional probability p on infinite intervals, there exists a real number $q > 0$ such that for $c \ge a$,*

$$p_{[a,\infty)}([c,d]) = \left(\frac{c}{a}\right)^{-q} - \left(\frac{d}{a}\right)^{-q},$$

and in general,

$$p_{[a,\infty)}([c,d]) = p_{[a,\infty)}([c,d] \cap [a,\infty)).$$

We thus have a 1-parametric family of probability distributions, that depends on a parameter q. For a finite subinterval, the distribution should be approximately uniform, so we expect the value of q to be small. When q is small, we can simplify the expression for probabilities by expanding this expression into Taylor series and keeping only linear terms in this expression. For an exponential function, this leads to $a^{-q} \approx 1 - q \cdot \ln(a)$ and, therefore,

$$p_{[a,\infty)}([c,d]) \approx (1 - q \cdot \ln(c/a)) - (1 - q \cdot \ln(d/a)) = q \cdot (\ln(d) - \ln(c)).$$

This formula is very similar to Benford's law. Indeed, Benford's law can be thus explained:

Proposition 9. *Assume that p is a subjective 1-dimensional probability p on infinite intervals, and that we know that $x \in [1,\infty)$. Then, the probability that the first digit in the decimal representation of x is d is equal to*

$$\frac{d^{-q} - (d+1)^{-q}}{1 - 10^{-q}}.$$

When $q \to 0$, this probability tends to $\log_{10}(d+1) - \log_{10}(d)$.

In the Proofs section, we provide the proof of this result. In addition to that proof, we want to give its intuitive explanation. Namely, let's compute the conditional probability of x having a leading digit d under the condition that $1 \le x \le 10^N$. Numbers with leading digit d belong to the intervals

$$[d, d+1) \cup [d \cdot 10, (d+1) \cdot 10) \cup \ldots$$

The total probability of belonging to these intervals is equal to the sum of the probabilities of belonging to $[d, d+1)$, to $[d \cdot 10, (d+1) \cdot 10)$, etc. Each of these probabilities is equal to $q \cdot (\ln(d+1) - \ln(d))$, and there are N of them, so we get $k \cdot q \cdot (\ln(d+1) - \ln(d))$. To get the desired conditional probability, we must divide this probability by the probability that $x \le 10^N$, which is $q \cdot \ln(10^N) = q \cdot N \cdot \ln(10)$. After division, we get the desired formula.

8 Applications: Case of Infinite Intervals

In addition to above mentioned accounting applications, Benford's law is used:

- in the design of (pseudo-)random number generators (see, e.g., [17]);
- for *comparing* different *roundings* in computer arithmetic, so that we can choose the rounding algorithm for which the average error (in the sense of the empirical distribution) is the smallest [8];
- a new *computer representation of real numbers* has been designed, that decreases the average rounding errors ("average" in the sense of this empirical distribution) [6, 30, 31]. This representation is called a *sli* (*s*ymmetric *l*evel *i*ndex) arithmetic, and it is defined as follows: for integers x, we define $\phi(x)$ as follows: $\phi(0) = 0$, $\phi(x+1) = \exp(\phi(x))$ (so that $\phi(1) = e$, $\phi(2) = e^e$, etc). This function ϕ is extended to a function that is defined for all real numbers and maps R to $[1, \infty)$. So, a number ≥ 1 can be represented as $\phi(r)$ for some r. Then, an arbitrary real number x is represented as a triple consisting of a rational number r and two signs, for which $x = \pm\phi(r)^{\pm 1}$. An interval is represented as by its upper and lower endpoints.

9 Multi-Dimensional Case: Infinite Intervals

Definition 5. *By an n-dimensional subjective probability for infinite intervals, we mean a function p that to every infinite box $B = [a_1, \infty) \times \ldots \times [a_n, \infty)$, $a_i > 0$, puts into correspondence a probability measure p_B on R^n that satisfies the following properties:*

- *For each box B, the value $p_B(C)$ is defined for every box $C = [c_1, d_1] \times \ldots \times [c_n, d_n]$ (finite or infinite).*
- *For every box B, the measure p_B is localized on the box B (i.e., $p_B(B) = 1$).*
- *The measures p_B that correspond to different sets B are consistent, i.e.,*

$$p_C(X) = \frac{p_B(X \cap C)}{p_B(C)}$$

  for arbitrary boxes B, C, and for an arbitrary set X.

- *The measures are unit-invariant, i.e.,*

$$p_B(X) = p_{\vec{\lambda} \cdot B}(\vec{\lambda} \cdot X)$$

  for an arbitrary vector $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$ with $\lambda_i > 0$.

Here, as in the case of finite intervals, unit-invariance means changing n measuring units.

Proposition 10. *If p is a subjective n-dimensional probability, then there exist positive real numbers $q_1, \ldots, q_n$ such that when $B = [a_1, \infty) \times \ldots \times [a_n, \infty)$,*

$$p_B([c_1, d_1] \times \ldots \times [c_n, d_n]) =$$

$$\left(\left(\frac{c_1}{a_1}\right)^{-q_1} - \left(\frac{d_1}{a_1}\right)^{-q_1}\right) \cdot \ldots \cdot \left(\left(\frac{c_n}{a_n}\right)^{-q_n} - \left(\frac{d_n}{a_n}\right)^{-q_n}\right).$$

10 Proofs

10.1 Proof of Proposition 1

For every $\alpha \in [0,1]$, let us denote $p_{[0,1]}([0,\alpha])$ by $f(\alpha)$. This value is defined for all $\alpha \in [0,1]$, because we assumed that all measures $p_{[a,b]}$ are defined for all the intervals $[c,d]$.

1°. Since $p_{[0,1]}$ is a probability measure, the function f is monotonically non-decreasing, and $f(0) = 0$.

2°. From the condition that the measure $p_{[a,b]}$ is localized on the interval $[a,b]$, we can conclude that $f(1) = 1$.

3°. As a particular case of a consistency requirement, we conclude that for arbitrary $\alpha \in [0,1]$ and $\beta \in [0,1]$, we have

$$p_{[0,\alpha]}([0, \alpha \cdot \beta]) = \frac{p_{[0,1]}([0, \alpha \cdot \beta])}{p_{[0,1]}([0, \alpha])}.$$

In our denotations, the right-hand side takes the form $f(\alpha \cdot \beta)/f(\alpha)$. Applying unit-invariance (with $\lambda = \beta$), we conclude that the left-hand side of this equality is equal to $p_{[0,1]}([0,\beta])$, i.e., in our notations, to $f(\beta)$. So, the above equality takes the form $f(\beta) = f(\alpha \cdot \beta)/f(\beta)$, or $f(\alpha \cdot \beta) = f(\alpha) f(\beta)$. In other words, we get a functional equation for the function f.

4°. This particular functional equation is well known: it has been first solved in [26], and its most general monotonic solution is [1], Section 3.1.1: $f(\alpha) = \alpha^q$ for some real number q.

To complete the proof of the theorem, we must show that $q = 1$.

5°. To do that, let us now consider another particular case of the consistency requirement:

$$p_{[0.5,1]}([0.5, 0.75]) = \frac{p_{[0,1]}([0.5, 0.75])}{p_{[0,1]}([0.5, 1])}.$$

5.1°. Let's first process the right-hand side. Since p is a probability measure, we have $p_{[0,1]}([0.5, 0.75]) = p_{[0,1]}([0, 0.75]) - p_{[0,1]}([0, 0.5]) = f(0.75) - f(0.5) = 0.75^q - 0.5^q$. Similarly, $p_{[0,1]}([0.5, 1]) = 1^q - 0.5^q = 1 - 0.5^q$. So, the right-hand side takes the form $(0.75^q - 0.5^q)/(1 - 0.5^q)$.

5.2°. Due to shift-invariance (with $c = 0.5$), the left-hand side of the equality from 5°. can be proven to be equal to $p_{[0,0.5]}([0, 0.25])$. Applying unit-invariance with $\lambda = 0.5$, we can now conclude that this expression is equal to $p_{[0,1]}([0, 0.5]) = f(0.5) = 0.5^q$.

6°. Substituting the expressions from 5.1° and 5.2° instead of the left- and the right-hand sides the equality from 5°, we conclude that

$$0.5^q = \frac{0.75^q - 0.5^q}{1 - 0.5^q}.$$

Multiplying both sides of this equality by the denominator, we conclude that $0.5^q - 0.25^q = 0.75^q - 0.5^q$. If we move each negative term to the opposite side of this equality, we will conclude that $2 \cdot 0.5^q = 0.25^q + 0.75^q$. Dividing the resulting equality by 2, we get

$$0.5^q = \frac{0.25^q + 0.75^q}{2}.$$

7°. There are three possibilities for q: $q > 1$, $q < 1$, and $q = 1$. To prove that $q = 1$, we must prove that the first two cases are impossible.

7.1°. For $q > 1$, the function $x \to x^q$ is strictly convex (because $(x^q)'' = q(q-1)x^{q-2} > 0$). Since $0.5 = (0.25 + 0.75)/2$, for $q > 1$, we will have

$$0.5^q < \frac{0.25^q + 0.75^q}{2}.$$

So, we cannot have $q > 1$.

7.2°. Similarly, when $q < 1$, we have $(x^q)'' = q(q-1)x^{q-2} < 0$, so the function $x \to x^q$ is concave and therefore,

$$0.5^q > \frac{0.25^q + 0.75^q}{2}.$$

So, the case $q < 1$ is also impossible.

7.3°. So, the only possible value is $q = 1$.

8°. Hence, $f(\alpha) = \alpha^q = \alpha$. In other words, for every $c \in [0, 1]$, $p_{[0,1]}([0, c]) = c$. Therefore, for every interval $[c, d] \subseteq [0, 1]$, we have $p_{[0,1]}([c, d]) = p_{[0,1]}([0, d]) - p_{[0,1]}([0, c]) = d - c = |[c, d]|$. The measure $p_{[0,1]}$ is localized on $[0, 1]$, therefore, the measure of any other other interval $[c,d]$ is determined only by this interval's intersection with [0,1]. Hence, for an arbitrary interval $[c,d]$, we have

$$p_{[0,1]}([c,d]) = |[0,1] \cap [c,d]| = \frac{|[0,1] \cap [c,d]|}{|[0,1]|}.$$

So, we have proved the desired formula for $[a,b] = [0,1]$.

9°. The formula for other intervals $[a,b]$ follows from the one that we have just proved, if we apply shift- and unit-invariance to transform [0,1] into $[a,b]$ (namely, we first apply a shift with $c = a$, and then $x \to \lambda \cdot x$ with $\lambda = b - a$). Q.E.D.

10.2 Proof of Proposition 2

1°. If we fix the component intervals of all the variables but one at [0,1], we get the 1-dimensional measure that satisfies all the conditions of Proposition 1. Therefore, from Proposition 1, we conclude that

$$p_I([0,1] \times \ldots \times [0,1] \times [c_i,d_i] \times [0,1] \times [0,1]) = d_i - c_i,$$

where by I, we denoted the unit box $I = [0,1] \times \ldots \times [0,1]$.

2°. Let's illustrate the remaining part of the proof on the example of $n = 2$ (for $n > 2$, the proof is quite similar). Suppose that we have two intervals $[c_i, d_i]$, and we want to find an expression for $\mu = p_I(B)$ for a box $B = [c_1,d_1] \times [c_2,d_2]$. Due to consistency, we have $\mu_r = \mu/\mu_d$, where we denoted

$$\mu_r = p_{[0,1]\times[c_2,d_2]}([c_1,d_1] \times [c_2,d_2])$$

and

$$\mu_d = p_I([0,1] \times [c_2,d_2]).$$

Because of 1°, we have $\mu_d = d_2 - c_2$. To compute μ_r, we must apply unit-invariance with $\lambda_2 = d_2 - c_2$ and $\lambda_1 = 1$. This application leads to

$$\mu_r = p_I([c_1,d_1] \times [0,1]),$$

and due to 1°, to $\mu_r = d_1 - c_1$. Since $\mu_r = \mu/\mu_d$, we have $\mu = \mu_r \cdot \mu_d = (d_1 - c_1) \cdot (d_2 - c_2) = |C|$. Q.E.D.

10.3 Proof of Propositions 3–5

Let us first compute the desired probabilities, i.e., prove Proposition 4.

i) The interval $\mathbf{a} - \mathbf{b}$ has the form $[\underline{a} - \overline{b}, \overline{a} - \underline{b}]$. Therefore, according to Proposition 1, the probability $p_{\mathbf{a}-\mathbf{b}}[0,\infty)$ is either equal to 0 (if $\overline{a} - \underline{b} < 0$), or, if $\overline{a} \geq \underline{b}$, to $|[0, \overline{a} - \underline{b}]|/|[\underline{a} - \overline{b}, \overline{a} - \underline{b}]| = (\overline{a} - \underline{b})/(\overline{a} - \underline{b} - \underline{a} + \overline{b})$.

ii) ([32]) According to Proposition 2, 2-dimensional subjective probability is proportional to the area. So, the desired probability is equal to the fraction whose denominator is the area $(\overline{a}-\underline{a})(\overline{b}-\underline{b})$ of the rectangle $[\underline{a},\overline{a}]\times[\underline{b},\overline{b}]$, and the numerator is the area of the portion of that rectangle for which $a \geq b$. This area is bordered by the line $a = b$, and by the sides of the rectangle. The area consists of three parts (some of which may be absent):

- The triangular part that comes from the intersection $[\max(\underline{a},\underline{b}),\min(\overline{a},\overline{b})]$ of the two intervals; its area is exactly the half of the area of the square (if there is an intersection at all).

- The area in which $a \geq \overline{b}$. In this area, all values a are $\geq \overline{b}$ and therefore, $\geq b$. Geometrically, this area is a rectangle $[\overline{b},\overline{a}]\times[\underline{b},\overline{b}]$ (if it exists at all, i.e., if $\overline{a} \geq \overline{b}$), so its area is equal to $(\overline{b}-\underline{b})\cdot\max(0,\overline{a}-\overline{b})$.

- The area in which $b \leq \underline{a}$. In this area, all values a are $\geq \underline{a}$ and therefore, $\geq b$. Geometrically, this area is a rectangle $[\underline{b},\underline{a}]\times[\underline{a},\overline{a}]$ (if it exists at all, i.e., if $\underline{b} \leq \underline{a}$), so its area is equal to $(\overline{a}-\underline{a})\cdot\max(0,\underline{a}-\underline{b})$.

Adding up these three areas and dividing by $(\overline{a}-\underline{a})(\overline{b}-\underline{b})$, we get the desired formula.

Now, we are ready to prove Proposition 3.

i) $\leftrightarrow$ iii). Due to part i) of Proposition 4, if $\overline{a} \geq \underline{b}$, then $p_{\mathbf{a}-\mathbf{b}}[0,\infty) \geq 1/2$ if and only if $(\overline{a}-\underline{b})/(\overline{a}-\underline{b}-\underline{a}+\overline{b}) \geq (1/2)$. Multiplying both sides by both denominators, we can conclude that $2\overline{a}-2\underline{b} \geq \overline{a}-\underline{b}-\underline{a}+\overline{b}$. Moving all negative terms to the other side of this inequality, we get the desired equivalent form $\overline{a}+\underline{a} \geq \overline{b}+\underline{b}$.

This equivalence was proved only under the auxiliary condition $\overline{a} \geq \underline{b}$. If $p \geq 1/2$, then this condition is satisfied (else, the probability is 0). So, to complete the proof of this equivalence, it is necessary to proof that if $\overline{a}+\underline{a} \geq \overline{b}+\underline{b}$, then $\overline{a} < \underline{b}$ is impossible. Indeed, if the upper bound $\overline{a}$ of an interval $\mathbf{a}$ is smaller than the lower bound $\underline{b}$ of the interval $\mathbf{b}$, then the midpoint $(\underline{a}+\overline{a})/2$ of $\mathbf{a}$ is definitely smaller than the midpoint $(\underline{b}+\overline{b})/2$ of $\mathbf{b}$, which contradicts to our assumption that $\overline{a}+\underline{a} \geq \overline{b}+\underline{b}$. The equivalence is thus proved.

ii) $\leftrightarrow$ iii). Let us use formula ii) from Proposition 3 to prove that this equivalence holds for all possible mutual locations of the intervals $\mathbf{a}$ and $\mathbf{b}$:

- If $\underline{a} \leq \overline{b}$, then the probability p (as computed by formula ii)) is 1 (which is $> 1/2$), and the midpoint of $\mathbf{a}$ is evidently larger than the midpoint of $\mathbf{b}$.

- If $\overline{a} \leq \underline{b}$, then $p = 0$ (i.e., $< 1/2$), and the midpoint of $\mathbf{a}$ is smaller than the midpoint of $\mathbf{b}$.

- If $\underline{a} \le \overline{a}$ and $\underline{b} \le \overline{b}$, then out of three terms in the numerator of of formula ii) only one term remains, so $p = (1/2)(\overline{a}-\underline{b})^2/(\overline{a}-\underline{a})(\overline{b}-\underline{b})$. So, $p \ge 1/2$ if and only if $(\overline{a}-\underline{b})^2 \ge (\overline{a}-\underline{a})(\overline{b}-\underline{b})$. But the length $\overline{a}-\underline{b}$ of the intersection is not greater than the length of each interval; so, $\overline{a}-\underline{b} \le \overline{a}-\underline{a}$, $\overline{a}-\underline{b} \le \overline{b}-\underline{b}$, and, therefore, $(\overline{a}-\underline{b})^2 \le (\overline{a}-\underline{a})(\overline{b}-\underline{b})$. If the intersection is actually smaller than one of the intervals, then we get the strict inequality. So, in this case, the only possibility for the left-hand side can be greater than or equal to the right hand side is when the intersection is actually equal to both intervals, i.e., when these two intervals coincide. In this case, the sums $\underline{a}+\overline{a}$ and $\underline{b}+\overline{b}$ coincide. If one of the ends is smaller, then $p < 1/2$ and $\underline{a}+\overline{a} < \underline{b}+\overline{b}$. So, in this case, $p \ge 1/2$.
- If $\underline{a} \ge \overline{a}$ and $\underline{b} \ge \overline{b}$, then:
  - on the one hand, $\underline{a}+\overline{a} \ge \underline{b}+\overline{b}$, and,
  - on the other hand, similarly to the previous case, $P(b > a) \le 1/2$ and therefore, $P(a \ge b) = 1 - P(a > b) \ge 1/2$.
- There are two remaining case: $\mathbf{b} \subseteq \mathbf{a}$ and $\mathbf{a} \subseteq \mathbf{b}$.
  - Let us first consider the first case, when $\underline{a} \le \underline{b} < \overline{b} \le \overline{a}$. In this case, due to the formula ii), the inequality $p \ge 1/2$ is equivalent to

    $$[(1/2)(\overline{b}-\underline{b})^2 + (\overline{b}-\underline{b})(\overline{a}-\overline{b})]/[(\overline{a}-\underline{a})(\overline{b}-\underline{b})] \ge 1/2.$$

    Multiplying both sides of this inequality by the denominators of both sides, we get the following equivalent inequality:

    $$(\overline{b}-\underline{b})^2 + 2(\overline{b}-\underline{b})(\overline{a}-\overline{b}) \ge (\overline{a}-\underline{a})(\overline{b}-\underline{b}).$$

    Dividing both sides by the positive number $\overline{b}-\underline{b}$, we get the equivalent inequality $\overline{b}-\underline{b}+2(\overline{a}-\overline{b}) \ge \overline{a}-\underline{a}$. If we do the multiplication, and move all negative terms to another side, we finally get the desired equivalent inequality $\overline{a}+\underline{a} \ge \overline{b}+\underline{b}$.
  - The proof for the second case, when $\mathbf{a} \subseteq \mathbf{b}$, is similar.

So, in all cases, ii) is equivalent to iii). Proposition 3 and 4 are proven.

Proposition 5 follows directly from Proposition 4. Q.E.D.

10.4 Proof of Proposition 8

This proof is similar to the proof of Proposition 1. Let us denote $p_{[1,\infty)}([\alpha,\infty))$ by $f(\alpha)$, where $1 \le \alpha < \infty$. This value is defined for all α, because we assumed that all measures $p_{[a,\infty)}$ are defined for all the intervals $[c,d]$ (finite or infinite).

1°. Since $p_{[1,\infty)}$ is a probability measure, the function f is monotonically non-decreasing, and $f(\alpha) \to 0$ as $\alpha \to \infty$.

2°. From the condition that the measure $p_{[1,\infty)}$ is localized on the interval $[1,\infty)$, we can conclude that $f(1) = 1$.

3°. As a particular case of a consistency requirement, we conclude that for arbitrary $\alpha \geq 1$ and $\beta \geq 1$, we have

$$p_{[\alpha,\infty)}([\alpha \cdot \beta, \infty)) = \frac{p_{[1,\infty)}([\alpha \cdot \beta, \infty))}{p_{[1,\infty)}([\alpha, \infty))}.$$

In our denotations, the right-hand side takes the form $f(\alpha \cdot \beta)/f(\alpha)$. Applying unit-invariance (with $\lambda = \beta$), we conclude that the left-hand side of this equality is equal to $p_{[1,\infty)}([\beta, \infty))$, i.e., in our notations, to $f(\beta)$. So, the above equality takes the form $f(\beta) = f(\alpha \cdot \beta)/f(\beta)$.

4°. From the proof of Proposition 1, we already know that the general monotonic solution of this equation is $f(\alpha) = \alpha^{-q}$. Since f is monotonically decreasing, we have $q < 0$. So, if $c \geq 1$, then $p_{[1,\infty)}([c,\infty)) = c^{-q}$.

5°. We are interested in the values of probability for finite intervals $[c, d]$, so we must somehow describe a finite interval in terms of infinite ones. If we simply take a difference $[c,\infty) - [d,\infty)$, we will get the formula for a measure of a semi-open interval $[c, d)$:

$$p_{[1,\infty)}([c, d)) = p_{[1,\infty)}([c,\infty) - [d,\infty)) =$$

$$p_{[1,\infty)}([c,\infty)) - p_{[1,\infty)}([d,\infty)) = c^{-q} - d^{-q}.$$

The formula for the probability of a closed interval $[c, d]$ can be deduced from the fact that an interval $[c, d]$ can be represented as a limit of the monotonically decreasing sequence of intervals $[c, d + 1/k)$ with $k = 1, 2, \ldots$. Therefore,

$$p_{[1,\infty)}([c, d]) = \lim_{k\to\infty} p_{[1,\infty)}\left(\left[c, d + \frac{1}{k}\right)\right) =$$

$$\lim_{k\to\infty}\left(\left(d + \frac{1}{k}\right)^{-q} - c^{-q}\right) = d^{-q} - c^{-q}.$$

The general case follows from unit-invariance. Q.E.D.

10.5 Proof of Proposition 9

The set of all the numbers ≥ 1 whose first digit is d consists of the infinite union of the intervals

$$[d, d + 1) \cup [d \cdot 10, (d + 1) \cdot 10) \cup \ldots$$

The total probability of belonging to these intervals is equal to the sum of the probabilities of belonging to $[d, d+1)$, to $[d*10, (d+1)*10)$, etc. The probability of belonging to $[10^k * d, 10^k * (d+1))$ is equal to

$$(10^k \cdot d)^{-q} - (10^k \cdot (d+1))^{-q} = (10^k)^{-q} \cdot D = 10^{-k \cdot q} \cdot D = c^k \cdot D,$$

where we denoted $c \stackrel{\text{def}}{=} 10^{-q}$ and $D \stackrel{\text{def}}{=} d^{-q} - (d+1)^{-q}$. So, the sum of these probabilities is the sum of a geometric progression

$$D + D \cdot c + D \cdot c^2 + \ldots + D \cdot c^k + \ldots$$

This sum is well known and equal to $D/(1-c)$. This is exactly the desired formula.

When $q \to 0$, we can use L'Hopital rule to compute the limit. Q.E.D.

10.6 Proof of Proposition 10

This proof is similar to the proofs of Propositions 1, 2, and 9.

1°. If we fix the component intervals of all the variables but one at $[1, \infty)$, we get the 1-dimensional measure that satisfies all the conditions of Proposition 9. Therefore, from Proposition 9, we conclude that for some constants q_i,

$$p_I([1, \infty) \times \ldots \times [1, \infty) \times [c_i, \infty) \times [1, \infty) \times [1, \infty)) = c_i^{-q_i},$$

where by I, we denoted the simplest infinite box $I = [1, \infty) \times \ldots \times [1, \infty)$.

2°. Let's illustrate the remaining part of the proof on the example of $n = 2$ (for $n > 2$, the proof is quite similar). Suppose that we have two infinite intervals $[c_i, \infty)$, and we want to find an expression for $\mu = p_I(C)$ for a box $C = [c_1, \infty) \times [c_2, \infty)$. Due to consistency, we have $\mu_r = \mu/\mu_d$, where we denoted

$$\mu_r = p_{[1,\infty)\times[c_2,\infty)}([c_1, \infty) \times [c_2, \infty))$$

and

$$\mu_d = p_I([1, \infty) \times [c_2, \infty)).$$

Because of 1°, we have $\mu_d = c_2^{-q_2}$. To compute μ_r, we must apply unit-invariance with $\lambda_2 = c_2$ and $\lambda_1 = 1$. This application leads to

$$\mu_r = p_I([c_1, \infty) \times [1, \infty)),$$

and due to 1°, to $\mu_r = c_1^{-q_1}$. Since $\mu_r = \mu/\mu_d$, we have $\mu = \mu_r \cdot \mu_d = c_1^{-q_1} \cdot c_2^{-q_2}$.

3°. To get the formula for finite boxes, we must express a finite box in terms of infinite ones. Namely, a finite box $F = [c_1, d_1] \times [c_2, d_2]$ can be represented as an intersection of two semi-finite boxes $F = F_1 \cap F_2$, where $F_1 = [c_1, d_1] \times [c_2, \infty)$

and $F_2 = [c_1, \infty) \times [c_2, d_2]$. Since p is a probability measure, we have $p(F_1 \cap F_2) = p(F_1) + p(F_2) - p(F_1 \cup F_2)$. So, to compute $p(F)$, it is sufficient to be able to compute $p(F_1)$, $p(F_2)$, and the probability of the union $F_1 \cup F_2$.

3.1°. The union $F_1 \cup F_2$ of these two boxes is an infinite box $[c_1, \infty) \times [c_2, \infty)$ for which we already know the probabilities.

3.2°. The set F_1 can be represented as a difference between the two infinite boxes, so $p(F_1) = p([c_1, \infty) \times [c_2, \infty)) - p([c_2, \infty) \times [c_2, \infty))$ (strictly speaking, we need to use semi-open intervals, and use a limit procedure as in the proof of Proposition 9).

3.3°. From these expressions, we will get the desired formula for the probability measure. Q.E.D.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209 and grant NCC2-1232, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant numbers F49620-00-1-0365, and by NSF grants CDA-9522207, ERA-0112968 and 9710940 Mexico/Conacyt.

References

[1] J. Aczél, *Lectures on functional equations and their applications,* Academic Press, N.Y., London, 1966.

[2] F. Benford, "The law of anomalous numbers", *Proc. Am. Phil. Soc.*, 1938, Vol. 78, pp. 551–572.

[3] F.-I. Buchholz, W. Kessel, F. Melchert, "Noise power measurements and measurement uncertainties", *IEEE Transactions on Instrumentation and Measurement*, 1992, Vol. 41, No. 4, pp. 476–481.

[4] CIPM (Comité International des Poids et Mesures), *Recommendation INC-I (1980)*, In: P. Giacomo, *News from the BIPM*, Metrologia, 1981, Vol. 17, pp. 69–74.

[5] CIPM (Comité International des Poids et Mesures), *Recommendation I (CI-1981)*, In: P. Giacomo, *News from the BIPM*, Metrologia, 1982, Vol. 18, pp. 41–44.

[6] C. W. Clenshaw, F. W. J. Olver, and P. R. Turner, "Level-index arithmetic: an introductory survey", In: *Numerical analysis and parallel processing,*

Springer Lecture Notes in Mathematics, Vol. 1397, Springer-Verlag, N.Y., 1989, pp. 95–168.

[7] W. M. Dong, W. L. Chiang, and H. C. Shah, "Fuzzy information processing in seismic hazard analysis and decision making", *International Journal of Soil Dynamics and Earthquake Engineering*, 1987, Vol. 6, No. 4., pp. 220–226.

[8] A. Feldstein and R. H. Goodman, "Some aspects of floating point computation", In: *Numerical analysis and parallel processing*, Springer Lecture Notes in Mathematics, Vol. 1397, Springer-Verlag, N.Y., 1989, pp. 169–181.

[9] B. J. Flehinger, "On the probability that a random number has leading digit A", *American Mathematical Monthly*, 1966, Vol. 73, pp. 1056–1061.

[10] J. R. Gott III, "Implications of the Copernican principle for our future prospects", *Nature*, Vol. 363, 27 May 1993, pp. 315–319.

[11] R. W. Hamming, "On the distribution of numbers", *Bell Systems Technical Journal*, 1970, Vol. 49, pp. 1609–1625.

[12] T. Hill, "A statistical derivation of the significant-digit law", *Statistical Science*, 1996, Vol. 10, pp. 354–363.

[13] T. P. Hill, "The first digit phenomenon", *American Scientist*, 1998, Vol. 86, pp. 358–363.

[14] E. T. Jaynes, "Information theory and statistical mechanics", *Physics Review*, 1957, pp. 106–108.

[15] V. V. Kafarov, B. V. Palukh, and V. L. Perov, "Solving the problem of technical diagnostics of uninterrupted production using interval analysis", *Doklady AN SSSR*, 1990, No. 3(13), pp. 677–680 (in Russian).

[16] R. B. Kearfott and Vladik Kreinovich, "Where to Bisect a Box? A Theoretical Explanation of the Experimental Results", In: G. Alefeld and R. A. Trejo (eds.), *Interval Computations and its Applications to Reasoning Under Uncertainty, Knowledge Representation, and Control Theory. Proceedings of MEXICON'98, Workshop on Interval Computations, 4th World Congress on Expert Systems*, México City, México, 1998.

[17] D. E. Knuth, *The art of computer programming. Vol. 2. Semi numerical algorithms*, Addison-Welsey, Reading, MA, 1969.

[18] V. Kreinovich, "Maximum entropy and interval computations", *Reliable Computing*, 1996, Vol. 2, No. 1, pp. 63–79.

[19] V. Kreinovich and T. Csendes, "Theoretical Justification of a Heuristic Subbox Selection Criterion", *Central European Journal of Operations Research CEJOR*, 2001, Vol. 9, No. 3, pp. 255–265.

[20] V. Kreinovich, H. T. Nguyen, and E. A. Walker, "Maximum entropy (MaxEnt) method in expert systems and intelligent control: new possibilities and limitations", In: K. M. Hanson and R. N. Silver, Eds., *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1996, pp. 93–100.

[21] V. Kreinovich, S. A. Starks, and G. Mayer, "On a Theoretical Justification of The Choice of Epsilon-Inflation in PASCAL-XSC", *Reliable Computing*, 1997, Vol. 3, No. 4, pp. 437–452.

[22] K. Löffler, "Behandlung von Wertebereichen in Datenbanken" ("Handling intervals in database"), *Angewandte Informatik*, 1987, Vol. 29, No. 1, pp. 20–22 (in German, English summary).

[23] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[24] B. V. Palyukh, "Technical diagnostics of production using interval methods", In: *Proceedings of the International Conference on Interval and Stochastic Methods in Science and Engineering INTERVAL'92*, Moscow, 1992, Vol. 1, pp. 131–133 (in Russian; English abstract Vol. 2, p. 88).

[25] B. V. Palyukh, B. V. Vasilyov, V. L. Perov, "Application of interval mathematics for solving technical diagnostics tasks of non-stop manufacture in chemical industry", *Interval Computations*, 1991, Vol. 1, No. 1, pp. 99–104.

[26] J. V. Pexider, "Notiz uber Funktionaltheoreme", *Monastch. Math. Phys.*, 1903, Vol. 14, pp. 293–301.

[27] R. S. Pinkham, "On the distribution of first significant digits", *Annals of Math. Statistics*, 1961, Vol. 32, pp. 1223–1230.

[28] R. A. Raimi, "On the distribution of first significant digits", *American Mathematical Monthly*, 1969, Vol. 76, pp. 342–348.

[29] M. Schneider, E. Shnaider, and A. Kandel, "Applications of the negation operator in fuzzy production rules", International Journal of Fuzzy Sets and Systems, 1990, Vol. 34, pp. 293–299.

[30] P. R. Turner, "A software implementation of *sli* arithmetic", In: Ercegovac, Swartzlander (eds.), *Proceedings ARITH 9*, IEEE Computer Society, Washington, DC, 1989, pp. 18–24.

[31] P. R. Turner, "Will the 'real' arithmetic please stand up?", *Notices of the American Mathematical Society*, 1991, Vol. 38, No. 4, pp. 298–304.

[32] D. Wagman, M. Schneider, and E. Shnaider, "On the use of interval mathematics in fuzzy expert systems", *International Journal of Intelligent Systems*, 1994, Vol. 9, pp. 241–259.

[33] WECC (Western European Calibration Cooperation), *Guidelines for the expression of the uncertainty of measurements in calibrations*, Document 19–1990, 1990.