Physics Letters A 380 (2016) 1895–1899

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Ubiquity of Benford's law and emergence of the reciprocal distribution

J.L. Friar [a], T. Goldman [a], J. Pérez-Mercader [b,c,*]

[a] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States
[b] Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, United States
[c] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, United States

ARTICLE INFO

Article history:
Received 15 February 2016
Received in revised form 29 March 2016
Accepted 31 March 2016
Available online 7 April 2016
Communicated by C.R. Doering

Keywords:
Benford/Reciprocal distribution
Information/Communication theory
Iteration theory

ABSTRACT

We apply the Law of Total Probability to the construction of scale-invariant probability distribution functions (pdf's), and require that probability measures be dimensionless and unitless under a continuous change of scales. If the scale-change distribution function is scale invariant then the constructed distribution will also be scale invariant. Repeated application of this construction on an arbitrary set of (normalizable) pdf's results again in scale-invariant distributions. The invariant function of this procedure is given uniquely by the reciprocal distribution, suggesting a kind of universality. We separately demonstrate that the reciprocal distribution results uniquely from requiring maximum entropy for size-class distributions with uniform bin sizes.

1. Introduction

In 1881 [1] the astronomer and mathematician Simon Newcomb observed that the front pages of tables of logarithms were more worn than later pages. In other words mantissas corresponding to quantities that had a smaller first digit were more common than for quantities with a larger first digit. He argued that the distribution of "typical" mantissas was therefore logarithmic. The physicist Frank Benford [2] rediscovered this in 1938 and provided more detail, for which his name is now associated with this phenomenon.

By now it is well documented that the frequency of first digits D in the values of quantities randomly drawn from an "arbitrary" sample follows Benford's Law of Significant Digits, namely,

$$B_b(D) = \frac{\ln(1+D) - \ln(D)}{\ln(b)} = \int_D^{1+D} \frac{dx}{x \cdot \ln(b)}\,, \tag{1}$$

where b is the arbitrary base for the logarithms and is commonly taken to be 10. We note that the probability of first digit 1 for base 10 is $\log_{10}(2) \cong .30$, far exceeding that for a uniform distribution of digits. The rightmost expression in Eqn. (1) expresses Newcomb's and Benford's logarithmic distribution as the cumulative distribution function (cdf) based on the **reciprocal** probability distribution function (pdf), which has been normalized to 1. The pdf that underlies Benford's Law is therefore the reciprocal distribution, $r(x) \equiv c/x$, with normalization constant $c = 1/\ln b$ when the random variable x ranges between $1/b$ and 1. We note that Eqn. (1) is base invariant (*i.e.*, invariant under a common change in the base of the various logarithms) and that the reciprocal pdf is scale invariant (a function $f(x)$ is said to be scale invariant if $f(\lambda x) = \lambda^p \cdot f(x)$ for any $p \in \mathbb{C}$). In this work we will concentrate on the emergence of the reciprocal distribution under a variety of conditions. The invariant (or fixed-point) function of an iterative procedure applied to distribution functions that are invariant under a continuous change of scales will be shown to be the reciprocal distribution. Additionally, requiring maximum entropy for size-class distributions with uniformly distributed bin sizes leads to the same function.

Very relevant to the discussion above is T.P. Hill's proof in 1995 [3–6] that random samples chosen from random probability distributions are collectively described by the reciprocal distribution, which is the pdf for the logarithmic or Benford distribution. In Hill's words: "If distributions are selected at random (in any "unbiased" way) and random samples are then taken from each of these distributions the significant digits of the combined sample will converge to the logarithmic (Benford) distribution." Because of this, the latter has been appropriately characterized as "the distribution of distributions," as Hill's theorem is in some sense the obverse (counterpart) of the Central Limit Theorem for probability **distributions** with large numbers of samples.

Benford's Law has been found to hold in an extraordinary number and variety of phenomena in areas as diverse as physics [7–12], genomics [13], engineering [14] and among many others, forensic

* Corresponding author at: Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, United States.
E-mail addresses: friar@lanl.gov (J.L. Friar), tjgoldman@post.harvard.edu (T. Goldman), jperezmercader@fas.harvard.edu (J. Pérez-Mercader).

http://dx.doi.org/10.1016/j.physleta.2016.03.045
0375-9601/

accounting [15]. Recently the number of examples where it applies has been expanding rather rapidly.

In the 1960's the need for understanding the constraints imposed in computation by finite word length and its impact on round-off errors were behind the interest of many, including R. Hamming [16,17], in Benford's law.

Importantly, Hamming argued that **repeated** application of any of the four basic arithmetic operations (addition, subtraction, multiplication and division) to numbers leads to **results** whose distribution of leading floating-point digits approaches the logarithmic (Benford) distribution. Hamming further argued that if any one arithmetic operation involves a quantity **already** distributed according to the reciprocal distribution, $r(x)$, then the result of this and all subsequent operations will result in quantities whose pdf for the leading floating-point digits is the reciprocal distribution. Hamming called this property the "**persistence** of the reciprocal distribution" although a better word might be **contagiousness**, since contact with the reciprocal distribution at any point in a calculational chain modifies the remaining chain irrevocably.

In this paper we use elementary methods to explore the connection between Benford's law, Hill's theorem and the "contagiousness" property of the reciprocal distribution. We will demonstrate this by constructing a simple but comprehensive class of probability distributions that depends on a single random variable that is dimensionless and unitless under a continuous change of scales. This class depends on an underlying pdf that is arbitrary, and which can be sampled in a manner consistent with Hill's Theorem. We further generalize this into an iterative procedure whose invariant functions are shown uniquely to be the reciprocal distribution, and which demonstrate Hamming's "contagiousness". Uniqueness obtains because the arbitrary (or "random" in this sense) underlying pdf eliminates any particular solutions in the invariant functions and leaves only the general solution. Our procedure generalizes the work of Hamming [16], and to the best of our knowledge is both new and useful. We show alternatively by invoking maximum entropy for a size-class distribution function that the reciprocal distribution again obtains as the unique solution. We conclude by speculating on the universality and applications of these results, with particular emphasis on minimizing errors in computations of various types.

2. Results

2.1. Invariance under changes in units and the law of total probability

In most scientific applications a stochastic variable x is assigned to the random values of some physical quantity. This quantity carries either physical dimensions (*e.g.*, length or volume) or units (such as the number of base pairs in a genome). However, because it refers to probabilities, the probability measure $F(x)\cdot dx$ that characterizes x must be dimensionless and unitless.

Hence, in order to remove units or dimensions from the measure it is necessary to introduce a parameter that results both in a dimensionless and unitless stochastic variable, **as well as in a bona fide** probability measure. Calling this parameter σ, for a specific value of σ we can rescale the physical variable x into a dimensionless and unitless random variable by just replacing x with $z = x/\sigma$. (We also assume for simplicity that x is positive semi-definite.) Then we can always introduce a normalizable function g such that

$$F(x) = \frac{1}{\sigma}\, g\left(\frac{x}{\sigma}\right), \tag{2}$$

and that has the correct properties expected from a probability measure. In other words, we can use a parameter σ to remove units or dimensions from the probability measure.

Familiar examples of distributions of the type g are the uniform distribution, $g_u(z) = \theta(1-z)$, the Gaussian distribution, $g_G(z) = \frac{2}{\sqrt{\pi}}\exp(-z^2)$, and the exponential distribution, $g_e(z) = \exp(-z)$, all of which satisfy the normalization condition: $\int_0^\infty dz\, g(z) = 1$. Heaviside step functions can be used for those cases where $g(z)$ is only non-vanishing in an interval, such as $z = [a, b]$, as was done above for the uniform distribution.

But the units chosen to measure x are, of course, arbitrary. For example, if x is a length, the units could be meter, millimeter, Angstrom, or even fathom, furlong, league, etc. In other words, the choice of units is itself arbitrary [18] and we can think of σ as a random variable with a distribution function $h(\sigma)$. Thus the problem we must study involves the combination of two stochastic variables. We can conveniently remove the scale and avoid the issue of units by using the Law of Total Probability [19] to combine the distribution g with a distribution of scale choices to produce a distribution $G(x)$:

$$G(x) = \int_0^\infty d\sigma\, \frac{g(x\,|\,\sigma)}{\sigma}\, h(\sigma)\,, \tag{3}$$

where now $G(x)$ and $h(\sigma)$ are interpreted as the **marginal** probabilities for events x and σ, and $g(x\,|\,\sigma)$ represents the conditional probability for x given σ. This well known law captures the intuitively clear statement that the probability that event x occurs is determined by summing the product of the probabilities for any of its antecedents σ to happen, times the conditional probability that x happens, **given** that σ has already occurred. Convergence of the integral for small values of σ is not a problem for $x \neq 0$ if $g(z)$ vanishes sufficiently rapidly for large z. Normalizability of g is sufficient for our purposes. The probability distribution in Eqn. (3) is fairly general and will be our template for studying the conditions underlying the emergence of the reciprocal distribution.

2.2. The Law of Total Probability and its recursive application

Let us consider a $g(x|\sigma)$ that is invariant under changes in dimensions or units. That is, let us assume that

$$g(x|\sigma) \equiv g\,(x/\sigma)\,, \tag{4}$$

with a concomitant interpretation for $g(x/\sigma)$ in the terms described in the preceding paragraph (N.B. the difference between "|" and "/"). Changing the integration variable to $z \equiv x/\sigma$ in Eqn. (4) leads to the convenient form

$$G(x) = \int_0^\infty \frac{dz}{z}\, g(z)\, h(x/z)\,. \tag{5}$$

It is important to note a property of Eqn. (5) that is a consequence of its structure: the function $G(x)$ has an **exceptional form** if $h(\sigma)$ is a scale-invariant (and power-law) function. A scale-invariant function $h(x/z)$ must be a power of its argument, or $h(x/z) \propto (x/z)^{-s}$ for a power-law. Ignoring the (for now) irrelevant proportionality constant, we then have

$$G(x) = \frac{1}{x^s}\int_0^\infty dz\, z^{s-1}\, g(z) \tag{6}$$

for $h(\sigma) = 1/\sigma^s$. We note that the integral $\int_0^\infty dz\, z^{s-1}\, g(z) = \mathcal{M}_s(g)$ is a constant and the Mellin transform [20] of the function $g(z)$. This allows one to rewrite Eqn. (6) in the more compact form

$$G(x) = \frac{1}{x^s}\, \mathcal{M}_s(g)\,. \tag{7}$$