axioms

MDPI

Essay

The Code Underneath

Julio Rives

Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Paseo Senda del Rey, 7, 28040 Madrid, Spain; jrives@dia.uned.es

Abstract: An inverse-square probability mass function (PMF) is at the Newcomb–Benford law (NBL)'s root and ultimately at the origin of positional notation and conformality. $\Pr(Z) = (2Z)^{-2}$, where $Z \in \mathbb{Z}^+$. Under its tail, we find information as harmonic likelihood $\mathcal{L}([s,t)) = H_{t-1} - H_{s-1}$, where H_n is the nth harmonic number. The global $\mathbb{Q}$-NBL is $\Pr(b,q) = {}^{\mathcal{L}([q,q+1))}/_{\mathcal{L}([1,b))} = (qH_{b-1})^{-1}$, where b is the base and q is a quantum $(1 \leq q < b)$. Under its tail, we find information as logarithmic likelihood $\ell([i,j)) = \ln {}^{j}/_{i}$. The fiducial $\mathbb{R}$-NBL is $\Pr(r,d) = {}^{\ell([d,d+1))}/_{\ell([1,r))} = \log_r(1 + {}^{1}/_{d})$, where $r \ll b$ is the radix of a local complex system. The global Bayesian rule multiplies the correlation between two numbers, s and t, by a likelihood ratio that is the NBL probability of bucket $[s,t)$ relative to b's support. To encode the odds of quantum j against i locally, we multiply the prior odds ${}^{\Pr(b,j)}/_{\Pr(b,i)}$ by a likelihood ratio, which is the NBL probability of bin $[i,j)$ relative to r's support; the local Bayesian coding rule is $\tilde{o}(j:i|r) = {}^{i}/_{j} \log_r {}^{j}/_{i}$. The Bayesian rule to recode local data is $\tilde{o}(j:i|r') = \tilde{o}(j:i|r)^{\ln r/\ln r'}$. Global and local Bayesian data are elements of the algebraic field of "gap ratios", $\frac{A-B}{C-D}$. The cross-ratio, the central tool in conformal geometry, is a subclass of gap ratio. A one-dimensional coding source reflects the global Bayesian data of the harmonic external world, the annulus $\{x \in \mathbb{Q} | \, 1 \leq |x| < b\}$, into the local Bayesian data of its logarithmic coding space, the ball $\{x \in \mathbb{Q} | \, |x| < 1 - {}^{1}/_{b}\}$. The source's conformal encoding function is $y = \log_r(2x - 1)$, where x is the observed Euclidean distance to an object's position. The conformal decoding function is $x = \frac{1}{2}(1 + r^y)$. Both functions, unique under basic requirements, enable information- and granularity-invariant recursion to model the multiscale reality.

Keywords: inverse-square law (ISL); Newcomb–Benford law (NBL); positional notation (PN); harmt (harmonic unit of information); likelihood; global-local duality; Bayesian law; secretary problem; gap ratio; cross-ratio; coding source; conformability; granularity

MSC: 03D20; 11A63; 11A67; 12F99; 30C35; 30F45; 33B15; 60-08; 60A99; 62C10; 65E10; 68P30; 83-10; 93A13; 94A17

Academic Editor: Kevin H. Knuth

Received: 31 October 2024
Revised: 17 January 2025
Accepted: 24 January 2025
Published: 30 January 2025

Citation: Rives, J. The Code Underneath. *Axioms* **2025**, *14*, 106. https://doi.org/10.3390/axioms14020106

1. Introduction

1.1. About the Newcomb–Benford Law

After Simon Newcomb's public note [1] and Benford's statement [2] that "small things are more numerous than large things, and there is a tendency for the step between sizes to be equal to a fixed fraction of the last preceding phenomenon or event", many scientists [3] tried to explicate the strange high frequency of the micro in nature, the rarity of the macro, and the ebbing progression of the gaps in between.

Nature pivots on exponential powers. Benford underlined that "the geometric series has long been recognized as a common phenomenon in factual literature and in the ordinary affairs of life". Nevertheless, human functions are often arithmetic-centric. Will there be a natural coding system to convert these realms into one another, the observable into our inner world's models, and vice versa? In other words, does nature count on a conformal transformation mechanism [4]?

In modern terms, the Newcomb–Benford law (NBL) states that the first digits of randomly chosen original data typically outline a logarithmic curve in an impressive diversity of fields, regardless of their physical units. Equivalently, the law remarks that raw natural data usually belong to nearly scale-invariant geometric series. Among its manifestations, it is fascinating that linear coefficients represented by mathematical and physical constants [5] (e.g., proportionality parameters or scalar potentials) adhere to the law.

Although this scenario suggests that the NBL might account for an elementary principle, we have yet to clarify its origin, realize a theoretical basis, or encounter a convincing reason [6]. Berger [7] laments that "There is no known back-of-the-envelope argument, not even a heuristic one, that explains the appearance of Benford's law across the board in data that is pure or mixed, deterministic or stochastic, discrete or continuous-time, real-valued or multidimensional".

We claim a primordial probability inverse-square law (ISL) is at the NBL's root. This "canonical" probability mass function (PMF) has a double fundamental effect, namely, the NBL for the discrete (global and harmonic) and continuous (local and logarithmic) domains. We prefer to anticipate these three laws' properties and affiliated terminology in Table 1, indicating their scope and character, baseline set, physical incarnation, scale, formula, information function, cardinality, and how we will denominate the corresponding item, an item list, and an item range.

Table 1. Nature and terminology of the three foundational contexts, where Z is a nonzero integer, b is the global base, q is a quantum ($1 \leq q < b$), H_n is the nth harmonic number, r is the local radix, and d is a digit ($1 \leq d < r$). The entry "Physics" poses a conjecture.

***Property* ↓ / *Law* →**	**Canonical PMF**	**First NBL**	**Second NBL**
Scope	Mathematical	Global	Local
Character	Discrete	Discrete	Continuous
Baseline set	Natural, Integer	Rational	Real
Physics	Field	Potential	Entropy
Entity at origin	Indeterminate	Observer	Coding source
Scale	Linear	Harmonic	Logarithmic
Probability law	$(2Z)^{-2}$	$(qH_{b-1})^{-1}$	$\log_r(1+1/d)$
Information function		Digamma	Logarithm
Cardinality	Infinite	Base	Radix
Item	Number	Quantum	Digit
Item list $<\alpha,\beta,\ldots>$	String	Chain	Numeral
Item range $[\alpha,\beta)$	Interval	Bucket	Bin

Are these laws naturally predetermined probability distributions? We champion the view that the canonical PMF is a brute fact and therefore the global and local versions of NBL are inescapable. For one thing, their mode is one. This number is the base case for almost all proofs by mathematical induction, statistically the most probable cardinal of a natural set (e.g., one cosmos, one black hole at the center of a galaxy, one star ruling an orbital planetary system, one heart pumping a body's blood, one nucleus regulating the cellular activity, et cetera), and seed in the majority of recursive computational processes. We read in [8] that numbers close to the multiplicative unit are not preferably rooted in mathematics, but a simple glance at the table of constants in [9] points in the opposite direction. Small leading digits and, in general, small significands (mantissae) of coefficients and magnitudes are the most common in sciences, albeit, of course, we can find cardinalities of all sizes.

1.2. Content

This article's field of study is mathematical and computational physics, delving into philosophy, theoretical physics, information theory, probability theory, and number theory.

We have organized it as follows (see Figure 1). We first examine the challenges researchers historically faced in deducing NBL and the state of the art in this field. Afterward, we present a one-parameter inverse-square PMF for the natural numbers with positive probabilities summing to one, extensible to the integers, and diverging mean (no bias). Next, we deduce the fiducial NBL passing through the global NBL; this two-phase derivation clarifies why the tendency for the minor numbers revealed by the natural sciences can be regular only if we assume that an all-encompassing base exists.

Information theory [10] comes into play when we discover that information is prior to probability in the context of NBL. Likewise, a unit fraction is the harmonic likelihood of an elemental quantum gap, and a digit of a numeral written in PN is a bin that covers a proportion of the available logarithmic likelihood.

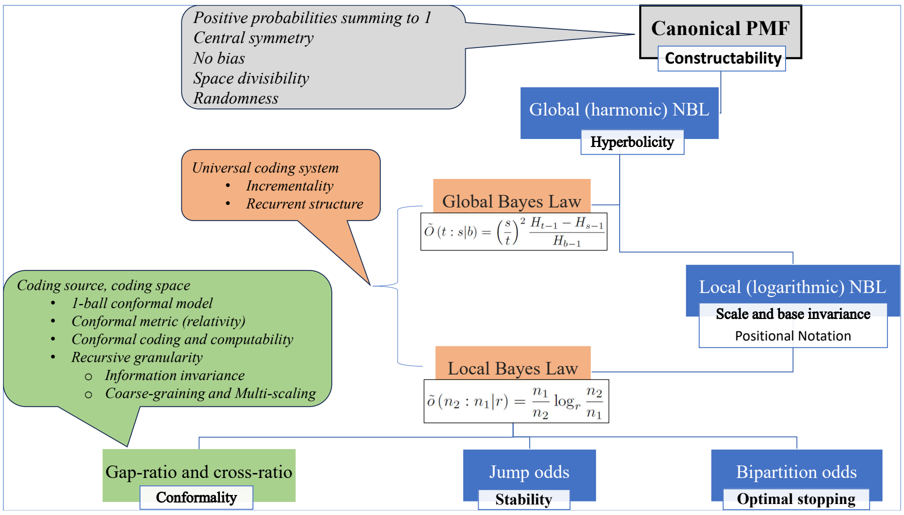

Figure 1. The road to conformal coding. We find information under the canonical PMF's tail as harmonic likelihood $\mathcal{L}([s,t)) = H_{t-1} - H_{s-1}$. The global $\mathbb{Q}$-NBL is $\Pr(b,q) = \mathcal{L}([q,q+1))/\mathcal{L}([1,b)) = (qH_{b-1})^{-1}$, where b is the base, and q is a quantum ($1 \leq q < b$). We find information under the $\mathbb{Q}$-NBL's tail as logarithmic likelihood $\ell([i,j)) = \ln j/i$. The fiducial $\mathbb{R}$-NBL is $\Pr(r,d) = \ell([d,d+1))/\ell([1,r)) = \log_r(1 + 1/d)$, where $r < b$ is the radix of a local complex system. The global Bayesian rule multiplies the prior correlation between numbers s and t, $\Pr(b,t)/\Pr(b,s) = (s/t)^2$, by a likelihood ratio that is the NBL probability of bucket $[s,t)$ relative to b's support. The local Bayesian rule multiplies the prior odds of quantum n_2 against n_1, $\Pr(r,n_2)/\Pr(r,n_1) = n_1/n_2$, by a likelihood ratio that is the NBL probability of bin $[n_1,n_2)$ relative to r's support. Jump and bipartition odds are problems where the synergy between Benford's and Bayes' rules manifests. However, we appreciate its full potential when a coding source, typically a complex system, conformally encodes the information perceived from the external (harmonic) world.

The odds between two events is a correlation measure whose entropic contribution to a positional scale ushers in Bayes' rule [11], namely, the product of two factors, a rational prior and a rational likelihood; precisely, the NBL probability of the numeric range involved. The structure of Bayesian data is recurrent under arithmetic operations and gives place to the algebraic field of "gap ratios". A subclass of gap ratio, the cross-ratio, is the central instrument of conformality and the ground for Lorentz covariance. Then, we determine the conformal metric and iterative coding functions that preserve the local Bayesian information and are compatible with a multiscale complex system [12].

1.3. Motivation and Method

Whereas the uniform distribution of probabilities is, in principle, fair and provides maximum entropy, it does not fit well into an open (infinite) domain. Contrariwise, it is noteworthy that [13] "the frequency with which objects occur in 'nature' is an inverse function of their size", indicating that oddity and magnitude usually correlate and conform to NBL. The cosmos displays a progressive aversion to larger and larger numbers, somewhat implementing the "parsimonae lex" [14], a principle of frugality that stimulates economy and effectiveness as universal prime movers, drivers of nascent physics.

That the universe is prone to favor slightness is particularly blatant in physics and chemistry. For instance, following the standard cosmological model [15], the abundance of hydrogen and helium is roughly 75% and 23% of all baryonic matter, respectively [16]. Higher atomic numbers than 26 (iron) are progressively more and more infrequent. Nevertheless, the universe's heaviest elements can comparatively produce the most remarkable galactic phenomena despite their shortage [17] (e.g., the necessary metals to form the Milky Way represent only about 2% of the galaxy's disk mass). Why do accessibility and reactivity maintain a hyperbolic relationship?

Consequently, the primary motivation of this work is seeking a reason for NBL and explaining its consequences rather than describing how it works [18] or elucidating its pervasiveness. If a genuine inverse-square PMF exists, we should arrive at it from just a few essentials. We confirmed that three preconditions, namely, positive probabilities summing to one, no bias, and central symmetry, unambiguously define a PMF, except for a proportionality constant. Moreover, requiring the average compartmentalization of the probability mass fixes such a constant and completely specifies the canonical PMF for the natural and integer numbers.

Because the resulting probability for counting numbers is a unit fraction, a rational version of NBL should accompany the logarithmic counterpart. $\mathbb{Q}$ fits into a pragmatic relational world ruled by proportions and approximations, contrasting with the physically unfeasible continuum's absolute density and ultra-accuracy of $\mathbb{R}$. Another motivation is disclosing how a coding source manipulates information in PN since NBL says nothing about the coding process that results in a digit's probability of occurrence. PN is the natural representation of a numeral on a coding source's logarithmic scale, enabling conformal transformations under certain constraints to exchange information between a system and its environment. The existing literature needs a theory that integrates Benford's and Bayes' laws at a fundamental level, not least leading to the concept of conformality.

1.4. Main Accomplishments

We have encountered that information has a relational character primally conveyed by the likelihood concept, either harmonic ($\mathcal{L}([s,t]) = H_{t-1} - H_{s-1}$harmt, i.e., "harmonic unit" of information) or logarithmic ($\ell([i,j]) = \ln j/i$nat). Likelihood is not the information obtained by picking an item from a range but the space allocated to encode an item between the range's ends. An NBL probability is a proportion of the information total (likelihood density), and an NBL entropy is the weighted mean of the information total (average likelihood). Moreso, odds, gap ratios, and cross-ratios measure likelihood correlations. Because algebra grows on these rational data, geometry embodies algebraic structures, and physics reflects geometrical rules, information turns out to be physical.

Another high-level achievement is finding a hidden connection between NBL and Bayes' law. This rudimentary rule codes the strength of the relationship between a pair of items normalized in a particular base b or radix r. The global Bayes rule, in odds form and b-ary harmonic information units, is the product of a prior, the ratio between the probability of two numbers t and s according to the canonical PMF, by a likelihood factor, the global NBL probability of the bucket $[s,t)$ in base b. The local Bayes rule, in odds form and r-ary logarithmic information units, is the product of the prior, a ratio between the global NBL probability of two quanta j and i on b's harmonic scale, by a likelihood factor, the local NBL probability of the bin $[i,j)$ in radix r. Further, Bayesian data conformally encoded constitute normalized likelihood information exhibiting as entropy. Bayes' rule also recodes information after a change of base or radix, a foundation for incremental computation. Lastly, we learned how a source recursively encodes the observable as Bayesian data and decodes these back into the information of the external world. This Bayesian outlook unifies the frequentist, subjective, likelihoodist, and information-theory interpretations.

This paper presents a theory of universal information coding that links mathematics with physics. It provides various instances of how a law, PMF, concept, or formula supports our theory that the cosmos is a hyperbolic, thrifty, and relational information system at a fundamental level. The notion of conformality implemented into a source's coding space subsumes these hallmarks. It employs the NBL invariance of scale and base in Bayes' rule to calculate the entropic contribution of a range of items. This synergy reinforces the thesis that mathematics begets physics and that information is a form of energy. The universe fosters a natural positional system that rules how a body's local quantum-mechanical degrees of freedom carve the information of its consubstantial properties, backing the Computable Universe Hypothesis [19].

1.5. Specific Achievements

We have found a roundabout but intuitive argument to explain the appearance of NBL in the vast array of contexts in which its effect manifests; NBL issues from an ISL of probability.

When choosing a natural number at random, nature follows a particular PMF, where zero is possible and interpretable as "indeterminate", e.g., not-a-number or inaction. We require zero's probability to be $1-\epsilon S$, where ϵ is a proportionality constant, and ϵS is the probability of picking a counting number, i.e., $\{1,2,3,\cdots\}$. We also need this one-parameter PMF to have no bias so that no number is prominent (up to its probability), i.e., any number can appear. Moreover, the mass of a counting number N is necessarily ϵ/N^2 if we want the probability function to be extensible to integer numbers, i.e., a number with the same probability regardless of the sign. Thus, the universe weighs the cost of choosing $\pm N$ as growing quadratically with N.

We have obtained the "global" and "local" NBL from this predetermined PMF. Under ϵ/N^2's tail, the probability that a natural number exceeds N is proportional to the trigamma function at N. Likewise, the probability of a natural variable's second-order cumulative function falling into $[s,t)$ is a harmonic likelihood ratio that cancels out the constant ϵ, namely, the bucket's width $\psi(t)-\psi(s)=H_{t-1}-H_{s-1}$ relative to the base's support width $\psi(b)-\psi(1)=H_{b-1}$, where $1\le s<t<b$ and $\{s,t,b\}\in\mathbb{N}^+$. The base b is a global referent that changes the status of a number to a computable elemental entity we call a "quantum". When the bucket is $[q,q+1)$, we obtain the global NBL of a generic quantum q, $\Pr(b,q)=\frac{1}{qH_{b-1}}$, an exact and separable function where $q,b\in\mathbb{N}$, $1\le q<b$, and H_n is the nth harmonic number. The global NBL represents, in information theory, the likelihood q encloses concerning the likelihood total; geometrically, a share of the surface area swept by q; and physically, a scalar potential harmonically diminishing as q moves away from the origin. The odds version of this PMF (16), also exact and separable, defines the stability of a quantum jump.

We can handle quanta as real variable values when the global base b is giant. Because a coding source does not know the value of b, it must establish a local referent $r \ll b$ to normalize its information separated from the surrounding environment, changing the status of a quantum to a locally computable elemental entity we call a "digit". This scenario involves the canonical PMF's third-order cumulative distribution; the probability of a quantum falling into $[i, j)$ is a logarithmic likelihood ratio that cancels out H_{b-1}, precisely the bin's width $\ln {}^{j}/_{i}$ relative to the radix support's width $\ln r$. When the bin is $[d, d+1)$, we arrive at the fiducial NBL, i.e., $\log_r(1 + {}^{1}/_{d})$, where $d \in \mathbb{N}$ is a digit such that $1 \leq d < r$. This PMF represents, in information theory, the likelihood d encloses regarding the likelihood r embraces. It is geometrically a hyperbolic sector equivalent to the surface area swept by d relative to that swept by r and physically a scalar potential r-logarithmically diminishing as d moves away from the origin.

In general, NBL probabilities consider the cutoffs PN imposes as a proportion of the total information. The global and local versions of NBL give probability masses similar to a degree. For comparison purposes, 1 in standard ternary occupies ${}^{2}/_{3} \approx 66.7\%$ ($\mathbb{Q}$) and ${}^{\ln 2}/_{\ln 3} \approx 63.1\%$ ($\mathbb{R}$), while 2 occupies ${}^{1}/_{3} \approx 33.3\%$ and ${}^{\ln 1.5}/_{\ln 3} \approx 36.9\%$, respectively. Likewise, 1 in standard decimal occupies 35.3% ($\mathbb{Q}$) and 30.1% ($\mathbb{R}$), while 9 occupies 3.9% and 4.6%, respectively.

NBL, a synonym of PN, a subsidiary of the canonical PMF, describes an information field where probability correlates with accessibility, whence with concentration and durability. Smaller significands occupy more room and furnish less information than greater significands. In other words, the space is denser and more stable near the coding source, while numerals dilute the space and become more reactive as we move away from the origin.

The analysis of NBL from the odds angle drives us to a rudimentary Bayesian framework. Bayesian encoding, recoding, and decoding are elemental computing routines that handle odds. The Bayesian encoding of the relation between two numbers is the entropic allocation of their correlation for a harmonic scale, i.e., their ratio squared multiplied by the probability of the associated interval in the chosen base (see Figure 2 in green). The Bayesian encoding of the relation between two quanta is the entropic contribution of their correlation for a logarithmic scale, i.e., their ratio multiplied by the probability of the associated bucket in the chosen radix (see Figure 2 in blue). This pattern allows us to interpret a Bayesian rule as the formula to encode the rational point ${}^{n}/_{d}$ or the corresponding range $[n, d)$ of integers; this duality principle asserting that points and lines are interchangeable is endemic to the cosmos.

The global Bayes law bridges numbers with information. We measure global Bayesian data in harmonic units of information that depend on the base. The natural harmonic scale uses bucket $[1, 2)$ as a reference. We measure local global Bayesian data in logarithmic units of information that depend on the radix. The natural logarithmic scale uses bin $[1, \mathrm{e})$ as a reference, where e is Euler's number.

The global Bayesian rule allows for calculating a quantum jump improbability, with masses decaying similarly to the global NBL as we move away from the source. Likewise, the local version of Bayesian coding drives us to the PMF of a domain's bipartition, an information function applicable to stopping problems. Specifically, we deduce the information gained from splitting a radix's digit set. If we take these digits as generic elements to be processed sequentially, our bipartite odds formula reaches a pair of information maxima involving e. The square root of the radix gives a minimum between the two maxima. We fix ideas by focusing on a variation of the "secretary problem" pursuing "a good" instead of "the optimal" solution. This problem's representativeness joins the overwhelming evidence supporting the overarching character of the NBL.

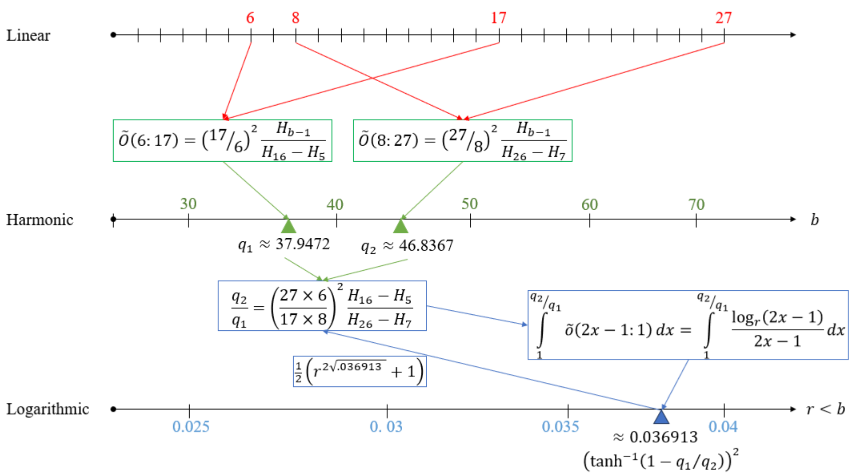

Figure 2. Example with $b = 101$ and $r = \mathrm{e}$ of how the three scales interact using Bayes' rules. q_2/q_1 is a global Bayesian datum corresponding to the local Bayesian datum $Q = 1 - q_1/q_2$; both represent the Euclidean distance to the origin. Within the logarithmic space of a coding source, the hyperbolic distance and the differential entropy from the origin to Q are $2\,\mathrm{artanh}(Q)$ and $\mathrm{artanh}^2(Q) \approx 0.036913$, respectively. The decoding function returns the latter value to q_2/q_1. Alternatively, the coding source can directly encode q_2/q_1 as $\log_r(2q_2/q_1 - 1) \approx 0.384255$ and decode it as $\frac{1}{2}(1 + r^{0.384255\ldots})$.

An exciting discovery is that the structure of Bayesian data is recurrent under arithmetic operations, giving rise to the algebraic field of gap ratios $\frac{A-B}{C-D}$. This set contains the cross-ratio, the linear fractional transformation's invariant over rings via the action of the modular group upon the real projective plane [20]. The logarithm of a cross-ratio locally provides us with the metric of a conformal space reflecting the observable world and consolidating the universal proclivity towards littleness, lightness, brevity, or shortness.

Restricted to one dimension, the cross-ratio's logarithm in radix r determines the coding space's metric with curvature $-\ln r$. The canonical encoding function $y = \log_r(2x - 1)$ and the canonical decoding function $x = \frac{1}{2}(1 + r^y)$ are the unique conformal transformations (i.e., preserving orientation and angles) that, if applied iteratively, map $x > 0$ to the coding space's positive side in accord with the minimal information principle. For the same reason, the hyperbolic distance $d_r(A, B) = 2/\ln r(\mathrm{artanh}(B) - \mathrm{artanh}(A))$ between points A and B inside the local coding space is also unique. We conclude that Poincaré invariance ultimately stems from the algebraic field of cross-ratios (with an infinite point).

The coding source expressly calculates the conformal distance from the origin as $2\,\mathrm{artanh}(Q)/\ln r$, where $\mathrm{artanh}(Q) = \sum_{i=1}^{\infty} \frac{Q^{2i-1}}{2i-1}$, $Q = \mathrm{sgn}(P) - 1/P$, $\mathrm{sgn}()$ is the sign function, and P is the observed Euclidean distance to the point where an external object is. The coding space is the ball $\{Q \in \mathbb{Q} |\, |Q| < 1 - 1/b\}$, with constant curvature of $-\ln r$, where r is the radix used to normalize the information; the harmonic (outside) and logarithmic (inside) scales have a common origin and are separated by the boundary ± 1 when $b \to \infty$. The "conformal encoding function" using the logarithm is $C = \mathrm{sgn}(P)\, \log_r(2\,\mathrm{sgn}(P)\, P - 1)$, with inverse "conformal decoding function" $\frac{1}{2}\,\mathrm{sgn}(C)\left(1 + r^{\mathrm{sgn}(C)\, C}\right)$. Since the metric ranges between $-\infty$ and ∞, the source can repeat the encoding process inwards until the external object's hyperbolic distance falls within the local coding space, halting the recursion. In the opposite direction, every 1-ball with a radius given by the iterated decoding of $C = 0$ outwards corresponds to a granularity level.

The results of this research all stem from the canonical PMF for the integer numbers, whose characteristics are fundamental and generative, imaging the essence of the cosmos. Physically, positive probabilities summing to one translates into unitarity, central reflection symmetry into parity invariance, fair mean and variance into uncertainty, holistic rationality into discreteness and relationalism, and utmost randomness in picking the number one into the principle of maximum entropy. Likewise, the global NBL (hence Zipf's law [21] with exponent 1), as well as the local NBL (supported by the logarithmic scale), has a tangible presence. More generally, our descriptions and derivations introduce diverse examples of how mathematical functions, rules, or algebraic structures emerge as observable dynamics.

2. The Whole Story of NBL

We comment on the aspects of the fiducial NBL that are most relevant to our research. Then, we introduce an inverse-square law as the origin of NBL. This PMF subsumes a probability law of rational masses, giving place to the harmonic scale, which employs a global base as a fundamental referent. When the global base is immense, the harmonic scale's rational setting approaches a domain of real variables and functions ruled by small radices in local environments. In other words, we prove that the local NBL, as everybody knows it, assumes that a prior all-encompassing base exists. Eventually, the interplay between the global base and the local radix will enable us to determine the canonical metric ascribed to a coding source's conformal space containing an image of the world.

2.1. The Road to the Fiducial NBL

The first digits of the numerals found in data series of the most varied sources of natural phenomena do not display a uniform distribution but rather we see that the minor ones are more likely (see [22] for a detailed bibliography and [23,24] for a general overview). Specifically, this "law of anomalous numbers" claims that the universe obeys an exponential distribution to a greater or lesser extent.

Newcomb's insight was: "The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally probable". (What Newcomb refers to as a "mantissa" is what we will call a "significand".) More than half a century later, Benford defined the exact formula of every random variable satisfying the first-digit (and other digits) law [2]. He could not derive it formally because mathematical difficulties arise in the transition from $\mathbb{N}$ to $\log \mathbb{N}$; when the baseline set is unlimited, it implies tackling the problem of "picking an integer at random" [25].

To commence, numerals beginning with a specific digit do not have a natural density. The decimal sequence $\{1, 11, 12, 13, \ldots, 100, 101, 102, 103 \ldots\}$ that groups the first digits does not converge (e.g., it oscillates). Moreover, suppose each natural occurs with equal probability. In that case, the whole space must have probability zero or infinity, violating countable additivity (by which the measure of a set must be nonzero, finite, and equal to the sum of the measures of the disjoint subsets); hence, we cannot construct a viable discrete probability distribution. The attempt to choose $P(N) = 1/N$ fails because it diverges in the limit and is not scale-invariant.

Hill [26] resumed Newcomb's idea; logarithm's significands of sequences conformant to NBL trace a uniform distribution. He identified an appropriate domain for the "natural probability space" and, based on the decimal mantissa σ-algebra (where countable unions and intersections of subsets can be assigned a gauge), formally deduced the law for the first digit and joint distribution of the leading digits. He also provided a new statistical log-limit central-limit-like significant-digit law theorem that stated the scale-invariance, base-invariance, sum-invariance, and uniqueness of NBL; for instance, mining data about the lifetime of mesons or antimesons in microseconds in decimal or seconds in binary results in strict observance of the law. Since Hill's publication in 1995, more derivations have come to light, one of the subtlest appearing in [27] (section 14.2). Nonetheless, they all ignore foundational causes.

A vehicle of NBL is how different measurement records spread and repositories aggregate data. For one thing, the significant-digit frequencies of random samples from random distributions converge to conform to NBL, even though some of the individual distributions selected may not [28]. Rarely does a distribution of distributions disagree with NBL [29]. The ratio distribution of two uniform, two exponential, and two half-normal distributions approximately stick to NBL. The Pareto distribution enjoys the scale-invariance property as long as we move from discrete to continuous variables, and Zipf's law ($\propto 1/z^\alpha$ with $\alpha \approx 1$) satisfies the above-mentioned absorptive property if one stays over the median number of digits [30]. More generally, right-tailed distributions putting most mass on small values of the random variable (i.e., survival or monotonically decreasing like the log-logistic distribution) are just about compliant with NBL [13] (e.g., the tail of the Yule–Simon distribution [31]). The log-normal distribution fits NBL, and the Weibull and inverse-gamma distributions are close to NBL under certain conditions [32]. Almost every exponentially increasing positive sequence is Benford (e.g., sequences of power a^n, where $a > 1$), and every super-exponentially increasing or decreasing positive sequence (e.g., the factorial) is Benford for almost every starting point [33]. In short, NBL embraces an ample range of statistical models and mixtures of probability distributions.

These mathematical circumstances explain why NBL is so widespread but not its reason. Failure to comprehend this distinction has generated confusion and is a typical scientific misunderstanding [34]. In other cases, authors have deemed specific remarks about NBL its cause when they are in fact consequences [35].

2.2. *A Fundamental Probability Law*

We seek a well-defined PMF, i.e., positive probabilities summing to 1. Not all Zipfian distributions [36] can do the job, for $\Pr(N) \propto N^{-a}$ eludes divergence only if $a > 1$. In general, linear forms for the denominator of a natural's probability cannot fulfill countable additivity.

We assume that $\mathbb{N}$ is an inductively constructible set from which all physical phenomena can crop up from the source outward, a basis of reductionism and weak emergency [37]. By including "nil", we also ponder "infinity" as its reciprocal. However, both projective concepts are only potential and limiting numbers in the offing; employing the successor and predecessor as symmetric constructors, we must be able to unrestrictedly choose any number strictly between 0 and ∞, so that no counting number is extraordinary. Again, many Zipfian distributions cannot do the job, for $\Pr(N) \propto N^{-a}$ has a diverging mean only if $a \leq 2$. For instance, cubic or higher polynomials lead to convergent expected values.

Additionally, we require a sound and dependable extension to the integers. Zipfian distributions where a is an even natural do the job, but in the range $a \in (1, 2]$ defined by the two previous requirements, $a = 2$ is the fitting choice, the only value assuring central reflection symmetry. To cap it all, $\Pr(N) \propto N^{-2}$ agrees with the minimal information principle [38]; considering other quadratic polynomials for the denominator of a natural's probability does not yield a better law because it would introduce unwarranted assumptions in vain. For instance, the Zipf–Mandelbrot law [39] $\Pr(N) = 1/(N^2 - 5N + 7)$ deals with unexplained coefficients and is not centrally symmetric.

Therefore, the PMF of a random variable X taking natural numbers is

$$\Pr(X = N) = \begin{cases} N \in \mathbb{N} - \{0\} : & \frac{\epsilon}{N^2} \\ else : & 1 - \epsilon\zeta(2) \end{cases} \quad (1)$$

We will suppose the proportionality parameter $\epsilon \in \mathbb{Q}^+$ to comply again with the minimal information principle. $\zeta(2) = \pi^2/6$ is the value of the Riemann zeta function at 2, brewing gently as a factor of endless aggregation of occurrence probabilities. Because the "else" (null) case is possible, this PMF is not a pure zeta distribution [40].

Countable additivity holds; the probabilities sum to 1 owing to

$$\sum_{N=1}^{\infty} \frac{\epsilon}{N^2} = \epsilon\zeta(2)$$

The picking event X is fair owing to the indeterminacy of the expected value of a natural number, i.e.,

$$\widehat{E}(X) = 0(1 - \epsilon\zeta(s)) + \sum_{N=1}^{\infty} N\frac{\epsilon}{N^2} = \epsilon\sum_{N=1}^{\infty}\frac{1}{N} = \infty$$

Indeed, the nth-order moment diverges for all nonzero $n \in \mathbb{N}$.

This PMF does not presume the law of large numbers or the law of rare events. On the contrary, it works under the statistical assumption of independence of occurrences and no bias. Outcomes of the picking event are unpredictable, even considering an indefinite trail of repetitions. No predetermined constant mean exists in space or time, nor is there an absolute measure of "rarity"; the relative frequency between two events solely depends on their probability mass. We can regard it as a brute law.

Let us leave the rational ϵ unfixed for the time being, given that it is unimportant for the derivation of NBL. Remember that $\epsilon \in (0, 6/\pi^2)$ holds the constraint $\Pr(N \in \mathbb{N}) > 0$ (i.e., $\epsilon/N^2 > 0$ and $\epsilon\zeta(2) < 1$), and we will return to it in Section 5.1.

2.3. The Rational (Global) Version of NBL

In analytic number theory, the mesmerizing Euler–Mascheroni constant γ ([41], section 1.5) is the limiting difference between the harmonic series and the logarithm, i.e.,

$$\lim_{N\to\infty} H_N - \ln N \sim \gamma$$

where $H_N \equiv \sum_{k=1}^{N} \frac{1}{k}$ is the Nth harmonic number. If our universe is as harmonic as logarithmic [27], the discrete version of the NBL must exist connected to but separated from the continuous (fiducial) one "at a distance of γ".

The cumulative distribution function of a random variable X obeying (1) is

$$\Pr(1 \leq X < N) = \epsilon\sum_{k=1}^{N-1}\frac{1}{k^2}$$

which tells us how often the random variable X is below N. We call its complementary function the "natural exceedance probability", quantifying how often X is on level N or above. This dwindling distribution function is

$$\Pr(X \geq N) = \Pr(X \geq 1) - \Pr(1 \leq X < N) = \epsilon(\zeta(2) - H_{N-1,2})$$

where $H_{N,2} \equiv \sum_{k=1}^{N} 1/k^2$ is the generalized Nth harmonic number in power 2.

We can express this probability in terms of the second derivative of the gamma function $\Gamma(x)$'s logarithm, i.e., the digamma function's first normal derivative, defined as

$$\psi'(x) = \left[\frac{\mathrm{d}}{\mathrm{d}x}\ln(\Gamma(x))\right]' = \frac{\mathrm{d}}{\mathrm{d}x}\frac{\Gamma'(x)}{\Gamma(x)}$$

Since $\zeta(2) = \psi'(1)$ and $\psi'(N+1) = \psi'(N) - 1/N^2 = \psi'(1) - H_{N,2}$, the natural exceedance of N is

$$\Pr(X \geq N) = \epsilon\psi'(N)$$

If numbers had physicality, the natural exceedance would represent a probability fractal signal with a probability density (per frequency range) that decays proportionally with the signal's frequency.

Regardless of the scale, let us divide the natural line into concatenated strings of numbers of the same length, which we name "quanta". Then, the second-order cumulative function arrives on the scene for global computability. The plot of $\epsilon\psi(q) + constant$, the natural exceedance's antiderivative, has an informational flavor. A significant value of the quantum q is more unpredictable and influential than a minor one; the only measure of its harmonic surprise is the extent of the event occurrence's log note $\psi(q)$.

So, how likely is the event $X = q$ to fall into bucket $[s, t)$, assuming a harmonic scale underneath? The "natural harmonic likelihood" $\mathcal{L}$ depends on the bucket's extent, namely,

$$[\mathcal{L}_X(q)]_s^t \equiv \frac{[\epsilon\psi(q) + constant]_s^t}{[\epsilon\psi(q) + constant]_1^2} = \frac{[\psi(q)]_s^t}{[\psi(q)]_1^2} = H_{t-1} - H_{s-1} = \mathcal{L}([s,t)) \in \mathbb{Q} \quad (2)$$

in natural harmonic units of (global) information, where we have considered the generalized recurrence relation $\psi(t+1) - \psi(s+1) = H_t - H_s$. Note that (2) is a proportion, canceling the constant ϵ. The natural harmonic likelihood is neither the probability of a quantum falling into $[s, t)$ nor the probability that $[s, t)$ is true given the observation $X = q$. It is the information attributable to the event $X = q$ when $s \le q < t$, i.e., the space allocated to encode a quantum between the bucket's ends, which is why it does not refer to q.

The harmonic number function (interpolated to cope with rational arguments) parallels the continuous world's logarithmic function in information theory, like in analytic number theory. The "harmt" (a portmanteau of "harmonic unit"), defined as

$$\mathcal{L}([1,2)) \equiv H_1 - H_0 = 1\text{harmt}$$

is the global unit of information, just as the natural local information unit, the "nat", corresponds to $[1, \mathrm{e})$ by $[\ln x]_1^{\mathrm{e}} = 1$. Thus, natural harmonic and logarithmic likelihoods are analogous. In particular, $\mathcal{L}[q, q+1) = \psi(q+1) - \psi(q) = H_q - H_{q-1} = 1/q$ implies that q's reciprocal denotes information, precisely the natural likelihood of an elemental quantum gap.

A global base b marks the boundary between the mathematical and physical world. We define the probability mass of bucket $[s, t)$ regarding b's support as the harmonic likelihood ratio

$$\Pr(b, [s,t)) = \frac{\mathcal{L}([s,t))}{\mathcal{L}([1,b))} = \frac{H_{t-1} - H_{s-1}}{H_{b-1}} \in \mathbb{Q} \quad (3)$$

where $1 \le s < t < b$ and $s, t, b \in \mathbb{N}$. This probability is separable as a product of $[s, t)$'s and b's functions, expressing a part of the information total that is the b-normalized rational quantum t/s's length or bucket $[s, t)$'s width.

The reader can object that the concept of likelihood is unnecessary to define (3) since we can directly define the probability of a bucket as $[\psi(q)]_s^t/[\psi(q)]_1^b$. However, we aim to stress that we obtain information, regardless of the base, only relative to the natural harmonic bucket $[\psi(q)]_1^2$. Because $[1, b)$ gives the maximum likelihood estimate, $\Pr(b, [s, t))$ is [42] the relative likelihood function of the bucket $[s, t)$ given $1 \le s < t < b$.

When $s = q$ and $t = q + 1$, we obtain

$$\Pr(b, q) = \frac{\mathcal{L}([q, q+1))}{\mathcal{L}([1,b))} = \frac{H_q - H_{q-1}}{H_{b-1}} = \frac{1}{H_{b-1}}\frac{1}{q} \in \mathbb{Q} \; (1 \le q < b) \quad (4)$$

We measure this PMF in b-ary harmonic information units. It is the simplest case of Zipf's law, geometrically an embryonic form of progressive one-dimensional circle inversion. Further, if q represented a frequency, we could understand the probability of a quantum with a given base as a (physical) potential diminishing hyperbolically with the distance from the source, i.e., a flicker [43] or pink [44] noise.

We have described how the harmonic series bridges Equations (1) and (4). Both laws point to minor numerical values as the most frequent significands, amassing more probability around the source to increase accessibility. However, we find three main differences between them:

1. Equation (1)'s probability masses are rational numbers. Instead, a quantum's probability represents an area ratio measured through the digamma function; hence, a quantum's probability is a quota of information.
2. The global NBL outlines a hyperbola instead of an ISL. Thus, while the probability of a number is inversely proportional to its norm (the number's square), the probability assigned to a quantum is inversely proportional to its modulus (the quantum's absolute value).
3. Equation (4) gives us the thing-as-it-appears (perceived potential) stemming from the thing-in-itself (field per se) [45] expressed by (1), two sides of the same property or object, the dual essence of the world.

2.4. Analysis of the Global NBL

The mean value for the quanta 1 to 9 following the global NBL is $q \approx 3.18$ (from $\Pr(10, q) = 1/9$), whereas it is $d \approx 3.43$ (from $\log(1 + 1/d) = 1/9$) for the local NBL. The harmonic mean value for the quanta 1 to 9 following the global NBL is $q = (9+1)/2 = 5$.

As expected, Equation (4) brings $\Pr(2, 1) = 1$, i.e., 1 occupies 100% of the space in binary. In base 3, the appearance probabilities of 1 and 2 as the first quantum are $\Pr(3, 1) = 2/3$ and $\Pr(3, 2) = 1/3$, respectively, a $2/1$ sharing out. We deem this Pareto rule so rudimentary that it might be fundamental in physics. The corresponding Pareto rule is $63/37$ if we utilize the local NBL. Quantum 1 in decimal occupies $\Pr(10, 1) = 2520/7129 \approx 35.3\%$, while it is 30.1% using the local NBL. Figure 3 compares the probability of a decimal datum's first position value between the global, discrete, rational, countable, harmonic NBL and the local, continuous, real, uncountable, logarithmic one. Regardless of the cardinality, the former is always steeper.

The unit bucket a quantum represents can be of any size, so we can recursively perform the integration and normalization process that gave rise to (4) "within" every quantum attributed to base b, obtaining a chain of nested quanta. The probability of obtaining the leading chain c of quanta with any length in b-ary is simply

$$\Pr(b, c) = \frac{H_c}{H_{b-1}} - \frac{H_{c-1}}{H_{b-1}} = \frac{1}{cH_{b-1}} \in \mathbb{Q}$$

This represents c's likelihood in b-ary harmonic units and becomes (4) when c is a base's quantum. For example, the probability masses that a decimal chain starts with "10" (e.g., 0.1071) and "99" (e.g., 992) are $\Pr(10, 10) = (10H_9)^{-1} \approx 0.03535$ and $\Pr(10, 99) = (99H_9)^{-1} \approx 0.00357$.

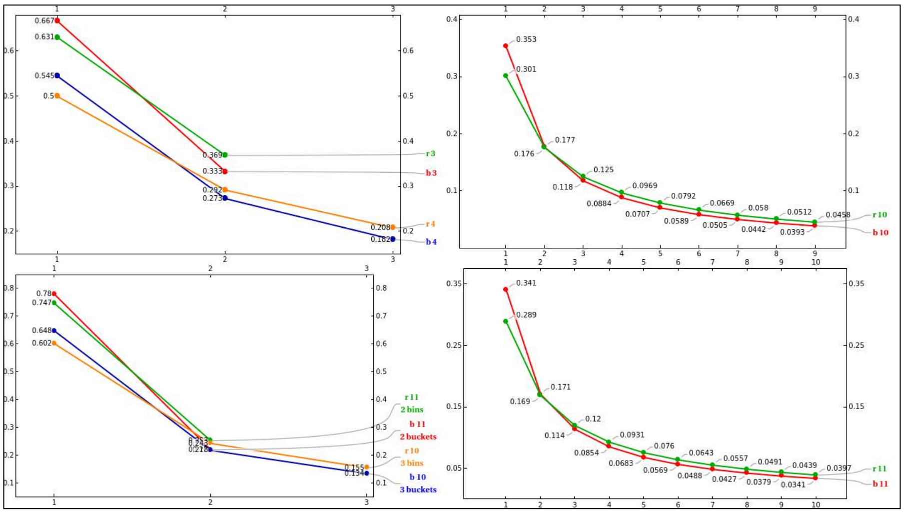

Figure 3. A comparison of the global with the local (fiducial) NBL, where "b" stands for "global base" and "r" for "local radix". Vertical axes represent the occurrence probability of the horizontal axes' quanta or digits. The plot on the top left shows the PMFs of the global and local standard ternary numeral system along with the PMFs of the global and local standard quaternary numeral system. The plot on the top right shows the PMFs of the global and local standard decimal numeral system. The plot on the bottom right shows the PMFs of the global and local standard undecimal numeral system. The plot on the bottom left shows the PMFs of the global and local standard undecimal numeral system divided into the partitions $[1,6)$ and $[6,11)$ and the PMFs of standard decimal divided into the partitions $[1,4)$, $[4,7)$, and $[7,10)$.

2.5. The Fiducial (Local) NBL

The global NBL furnishes the frame for constructing a sheer logarithmic system that conserves base and scale. To achieve such a pursuit, we must turn to the local context of a coding source and analyze how it represents a numeral in PN.

We call a "bin" of "digits" to a bucket of quanta in the source's proximity. The third-order cumulative function of (1) arrives on the scene to facilitate local computability. When the base b is enormous, we can handle digits like real values to calculate the antiderivative of (4), ${}^{\ln d}/_{H_{b-1}} + constant$, which outlines how unexpected and momentous digit d is. Large values locally transmit more information than small ones; for whom? Logarithmic surprise needs an observer to reify the event occurrence. The harmonic information perceived by a receiving system, a coding source, becomes local information with extension $\ln d$. Consequently, broad bins are more likely than narrow ones as supporting evidence.

Assuming a logarithmic scale underneath, we define the "natural logarithmic likelihood" ℓ_Y of the event $Y = d < b$ to fall into bin $[i, j)$ as the ratio

$$[\ell_Y(d)]_i^j \equiv \frac{\left[\frac{\ln d}{H_{b-1}} + constant\right]_i^j}{\left[\frac{\ln d}{H_{b-1}} + constant\right]_1^{\mathrm{e}}} = \frac{[\ln d]_i^j}{[\ln d]_1^{\mathrm{e}}} = \ln\frac{j}{i} = \ell([i,j)) \in \mathbb{R} \tag{5}$$

Note that this proportion no longer refers to base b; a coding source is unaware of the global setting for calculation purposes. The natural logarithmic likelihood is neither the probability of a digit falling into $[i, j)$ nor the probability that $[i, j)$ is true given the observation $Y = d$. It is the information attributable to the event $Y = d$ when $i \leq d < j$, i.e., the space allocated to encode a digit between the bin's ends, which is why it does not refer to d. However, it has nothing to do with surprisal [46]; ℓ denotes informative space rather than information content. Indeed, we can take it as the natural logarithmic length of ${}^{j}/_{i}$ or width of $[i, j)$. We can also take (5) as the differential entropy of the uniform probability density function $\Pr(x) = {}^{i}/_{j}$, where $0 < x < {}^{j}/_{i}$. We measure the natural logarithmic likelihood in natural units ("nats") because of $[\ln d]_1^{e} = 1$.

The domain of a digit d spans from the unit to $r - 1$, where $r \ll b$ is the cardinality of the local coding space, precisely the source's "radix". We define the r-ary probability mass of bin $[d_1, d_2)$ relative to the radix's support as the logarithmic likelihood ratio

$$\Pr(r, [d_1, d_2)) = \frac{\ell([d_1, d_2))}{\ell([1, r))} = \frac{\ln \frac{d_2}{d_1}}{\ln r} = \log_r \frac{d_2}{d_1} \in \mathbb{R}\ (1 \leq d_1 < d_2 < r) \tag{6}$$

with d_1, d_2, and $r \in \mathbb{N}$. We can take it as the representation length of d_2/d_1 or the width of $[d_1, d_2)$ in r-ary logarithmic information units, in correspondence with Equation (3), reckoning the probability of a bucket as a normalized harmonic likelihood. Therefore, in PN, the probability is a quota of the available space, a view we will develop in Sections 3.1 and 3.2.

Geometrically, the probability of event $d_1 \leq d < d_2$ conditioned to r is the ratio between the areas under the hyperbola delimited by bins $[d_1, d_2)$ and $[1, r)$, equivalent to the area enclosed by the rays $1/d_1$ and $1/d_2$ relative to the span of the hyperbolic angle r. Because the hyperbola preserves scale changes, the logarithm uniformly distributes the significant digits of a geometrical sequence, as Newcomb underlined in his note; $k \ln x = \ln x^k$ implies that, for example, x must drop to $\sqrt[3]{x}$ to divide the natural likelihood by three ($k = 1/3$).

By setting in (6) $d_1 = d$ and $d_2 = d + 1$, we fit Y's occurrences into the digits of a standard PN system with radix r, obtaining

$$\Pr(r, d) = \log_r \left(1 + \frac{1}{d}\right) \in \mathbb{R} \tag{7}$$

The original natural random variable $Y \in \mathbb{N}$ and the underlying global base b are absent. This expression is the local (fiducial) NBL, which tells us the PMF of an r-ary numeral's first digit.

A coding system (observer or source) that uses standard PN handles the unit range as a concatenation of the sub-bins $[\log_r 1, \log_r 2) = [0, \log_r 2)$, $[\log_r 2, \log_r 3)$, ... $[\log_r(r-1), \log_r r) = [\log_r(r-1), 1)$, covering intervals of $\log_r 2/1$, $\log_r 3/2$, ... $\log_r r/(r-1)$ units of space, and corresponding to the symbols 1, 2, ... and $r - 1$, respectively; the addition of these areas is the unit.

More fundamentally, common digits are near the coding source, i.e., the probability of a digit correlates with its accessibility and declines logarithmically. If we liken probability mass to space, smaller digits induce more density than significant digits. In other words, accessibility concentrated around the origin progressively dilutes as we move away, contrasting with the linear scale that distributes the space evenly.

We can generalize (6) to cope with bins outside the radix. The resulting expression is not generally a probability anymore, given that we can have bins of any size, but it is again an r-normalized likelihood that retains the geometric interpretation. In other words,

$$\ell([n_1, n_2)|r) = \frac{\ell([n_1, n_2))}{\ell([1, r))} = \log_r \frac{n_2}{n_1} \in \mathbb{R}\ (1 \leq n_1 < n_2) \tag{8}$$

is the r-normalized n_2/n_1's length or $[n_1, n_2)$'s width. We can regard it as a fractal dimension where r is the scaling factor, n_2 is the number of measurement units, and n_1 is the number of fractal copies. For instance, (8) might explain the Weber–Fechner law [47] in psychophysics, where $\ell([n_1, n_2)|r)$ is the intensity of human sensation, $1/\ln r$ is a perception- and stimulus-dependent proportionality constant, n_2 is the strength of the stimulus, and n_1 is the zeroing strength threshold.

When $n_1 = n$ and $n_2 = n + 1$, we can again interpret this likelihood as the probability of obtaining a leading r-ary numeral $n \in \mathbb{N}^+$ of any length, i.e.,

$$\Pr(r, n) = \ell([n, n+1)|r) = \log_r(n+1) - \log_r n = \log_r\left(1 + \frac{1}{n}\right) \in \mathbb{R}$$

The efficiency of an r-ary numeral system worsens as $r \to 1^+$ or $r \to b \to \infty$ [48] because r diverges from the optimal radix economy, namely, Euler's number e, destroying the information. In the former case, we encounter the unary system, which boils down to a linear frequency. In the latter case, the numerals $n < r$ that only use the first position increase limitlessly. Both are non-coding cases.

3. Odds

Although odds typically appear in gambling and statistics, this section illustrates how they are central to the computational processes of a coding source, including an application to physics and another to decision theory.

We usually define the odds of an outcome as the ratio of the number of events that generate that particular result to those that do not. In this sense, odds constitute another measure of the chance of a result that highlights its rational character. The ratio between the probabilities of two events determines their relative odds; the higher the odds of an outcome compared with another, the more informative the latter's occurrence is.

In Bayesian coding, odds between propensities or degrees of belief become information correlations representing entropic contributions to a system's hyperbolic scale. This stance embraces the objectivistic [49] and subjectivistic [50] interpretations.

3.1. Global Bayesian Coding

The probability ratio between two numerical events diverges from the unit as their correlation weakens. It fits into the base's harmonic scale of a global system multiplied by a likelihood factor that depends on the two events and the base. This operation is rigorously Bayes' theorem. Specifically, the formula

$$\tilde{O}(t : s|X = b) = \tilde{O}(t : s)\, \Lambda_{\tilde{O}}(t : s|X = b) \in \mathbb{Q}\ (1 \leq s < t < b) \tag{9}$$

encodes the relative odds between two numbers.

- $\tilde{O}(t : s|X = b)$ represents the global (encoded or posterior) odds of obtaining quantum t against s in base b. We can consider it the rational quantum s/t on a b-ary harmonic scale.
- $\tilde{O}(t : s)$ is the ratio between the probabilities of the two events according to (1), namely,

$$\tilde{O}(t : s) = \tilde{O}(t : s|1) = \frac{\Pr(t)}{\Pr(s)} = \left(\frac{s}{t}\right)^2$$

straightforwardly measuring the (decoded or prior) odds of picking the number t against s on a linear scale. If we fix the center of the range, the narrower the interval, the higher the odds, whereas if we fix the interval width, the minor s (or t), the lower the odds. Note that the odds of two concatenated intervals calculated separately are the product of the interval's odds:

$$\tilde{O}(z : x|1) = \left(\frac{x}{y}\right)^2 \left(\frac{y}{z}\right)^2 = \left(\frac{x}{z}\right)^2$$

- $\Lambda_{\tilde{O}}(t : s|X = b)$ is the global coding (Bayes) factor, which measures the degree to which the outcome b of the random variable X supports "hypothesis" t against s, assuming they are independent. Because interval $[s,t)$ is not yet encoded, the coding law establishes a likelihood difference instead of a likelihood ratio, namely,

$$\Lambda_{\tilde{O}}(t:s|X=b) = \mathcal{L}(t|X=b) - \mathcal{L}(s|X=b) = \frac{H_{t-1}-H_{s-1}}{H_{b-1}} = \Pr(b,[s,t))$$

where

$$\mathcal{L}(q|X=b) = \frac{H_{q-1}}{H_{b-1}}\ (1 \le q < b) \tag{10}$$

is the likelihood function of q with b fixed; since $\mathcal{L}(1|X=b)$ vanishes and $\mathcal{L}(b|X=b)=1$, we can understand this function as a measure, normalized to one, of the nearness between q and b. The coding factor is precisely Equation (3), measured in b-ary harmonic information units.

Compiling, PN calculates (9) as

$$\tilde{O}(t:s|b) = \left(\frac{s}{t}\right)^2 \frac{H_{t-1}-H_{s-1}}{H_{b-1}} \in \mathbb{Q}\ (1 \le s < t < b) \tag{11}$$

This is the cost of computing the bucket's harmonic width, i.e., the entropic contribution of bucket $[s,t)$ to b's harmonic scale. $\tilde{O}(2:1|b)$ is maximally informative, irrespective of the base. The global odds of a quantum against itself vanish, having no representation on a harmonic scale. The reciprocal

$$\tilde{O}(s:t|b) = \frac{1}{\tilde{O}(t:s|b)} = \left(\frac{t}{s}\right)^2 \frac{H_{b-1}}{H_{t-1}-H_{s-1}} \in \mathbb{Q}\ (1 \le s < t < b)$$

measures the odds of quantum s against t, with a maximum approaching b^2 as b climbs to infinity.

A global coding system must employ (9)'s variation

$$\tilde{O}\big(t:s|b'\big) = \tilde{O}(t:s|b)\,\Lambda_{\tilde{O}}\big(b':b|q \in [s,t)\big) \in \mathbb{Q}$$

to recode, where $1 \le s \le q < t < b$ and $t < b'$. The coding (Bayes) factor is a likelihood ratio when it deals with previously encoded data, as usual in statistics; using (10),

$$\Lambda_{\tilde{O}}\big(b':b|q \in [s,t)\big) = \frac{\mathcal{L}(b|q)}{\mathcal{L}(b'|q)} = \frac{H_{b-1}/H_{q-1}}{H_{b'-1}/H_{q-1}} = \frac{H_{b-1}}{H_{b'-1}}$$

measures the degree to which a given quantum supports "hypothesis" b' against b, assuming they are independent. This is the classical Bayes factor, replacing probabilities with global likelihoods. Hence, the global coding system can change to base b' utilizing the rule

$$\tilde{O}\big(t:s|b'\big) = \tilde{O}(t:s|b)\frac{H_{b-1}}{H_{b'-1}} \in \mathbb{Q} \tag{12}$$

which coincides with the odds of $[s,t)$ in base b' for the first time because of

$$\tilde{O}\big(t:s|b'\big) = \left(\left(\frac{s}{t}\right)^2\left(\frac{H_{t-1}-H_{s-1}}{H_{b-1}}\right)\right)\frac{H_{b-1}}{H_{b'-1}} = \left(\frac{s}{t}\right)^2 \Pr\big(b',[s,t)\big)$$

The transformation $b \to b'$ constitutes a primal memory (incremental) process that decreases the global odds if $b < b'$, and vice versa increases the global odds if $b > b'$. Imagine that a given universe region is a global coding system with an autonomous clock; then, b growing every tick would mean an unstoppable progressive information loss for that region. This connection between time and entropy is crucial to theoretical physics and cosmology [51].

Let us see some examples. Suppose a global coding system encodes bucket $[4,13)$ to base $b = 100$ as

$$\tilde{O}(13:4|100) = \left(\frac{4}{13}\right)^2 \frac{H_{13-1} - H_{4-1}}{H_{100-1}} \approx 0.0232$$

This value is the entropic contribution of bucket $[4,13)$ to 100's harmonic scale. When the system changes the base to $b' = 110$, using (12), it delivers

$$\tilde{O}(13:4|110) = \tilde{O}(13:4|100)\frac{H_{100-1}}{H_{110-1}} \approx 0.0228$$

meaning that the bucket's entropic contribution decreases. Then, changing to base $b' = 90$ yields

$$\tilde{O}(13:4|90) = \tilde{O}(13:4|110)\frac{H_{110-1}}{H_{90-1}} \approx 0.0237$$

i.e., the bucket's entropic contribution increases. Finally, the system decodes the odds with base 90 by solving the prior from (11), i.e.,

$$\tilde{O}(13:4) = \frac{\tilde{O}(13:4|90)}{\Pr(90,[4,13))} = \tilde{O}(13:4|90)\frac{H_{90-1}}{H_{13-1} - H_{4-1}} = \left(\frac{4}{13}\right)^2$$

3.2. Local Bayesian Coding

Local Bayesian coding assumes that (9), the global Bayesian law, governs the universe's information. We express the informational correlation between two numerals by multiplying their harmonic correlation by a coding factor, obtaining a point on a logarithmic scale. This operation is rigorously the Bayes theorem that settles down the basis of a conformal metric space. Specifically, local Bayesian coding employs the formula

$$\tilde{o}(n_2:n_1|Y=r) = \tilde{o}(n_2:n_1|b)\,\Lambda_{\tilde{o}}(n_2:n_1|Y=r) \in \mathbb{R} \tag{13}$$

to encode the probability ratio between n_1 and n_2, where $(1 \leq n_1 < n_2 < b) \wedge (r < b)$.

- $\tilde{o}(n_2:n_1|Y=r)$ represents the local (encoded or posterior) odds of obtaining n_2 against n_1 with radix r.
- $\tilde{o}(n_2:n_1|b)$ is the (prior) probability ratio between the two events only assuming that a global base exists; using (4),

$$\tilde{o}(n_2:n_1|b) = \frac{\Pr(b,n_2)}{\Pr(b,n_1)} = \frac{\frac{1}{n_2 H_{b-1}}}{\frac{1}{n_1 H_{b-1}}} = \frac{n_1}{n_2}$$

This measures the strength of the association between n_1 and n_2 on the harmonic scale provided by b.

- $\Lambda_{\tilde{o}}(n_2:n_1|Y=r)$ is the local coding (Bayes) factor, which measures the degree to which the outcome r of the random variable Y supports "hypothesis" n_2 against n_1, assuming they are independent. Because the bucket $[n_1,n_2)$ is not locally encoded yet, the coding law establishes a likelihood difference instead of a likelihood ratio, namely,

$$\Lambda_{\tilde{o}}(n_2:n_1|Y=r) = \ell(n_2|Y=r) - \ell(n_1|Y=r)$$

where $\ell(n|Y=r) = \log_r n$ quantifies the likelihood of n when the actual value r occurs. In short, the local coding factor is the log-odds of n_2 relative to n_1, equivalent to the support of $[n_1,n_2)$ with radix r according to (8), i.e.,

$$\Lambda_{\tilde{o}}(n_2:n_1|Y=r) = \log_r n_2 - \log_r n_1 = \log_r \frac{n_2}{n_1} = \ell([n_1,n_2)|r)$$

Compiling, the PN system calculates (13) as

$$\tilde{o}(n_2 : n_1|r) = \frac{n_1}{n_2} \log_r \frac{n_2}{n_1} \ (1 \leq n_1 < n_2) \tag{14}$$

Consequently, the local Bayes' rule measures the entropic contribution of bin $[n_1, n_2)$ on r's logarithmic scale, with a minimum approaching $\tilde{o}(r : 1|r) = 1/r$ as r climbs to infinity. The local odds of a numeral against itself vanish, having no representation on a logarithmic scale. Euler's number has an extraordinary meaning in this setting; a Bayesian datum in this form is maximally informative irrespective of the radix when $n_2/n_1 = \mathrm{e} = 2.718\ldots$, an ideal proportion that induces the nat, the local information unit associated with bin $[\ln n]_1^{\mathrm{e}}$ we use in (5).

A PN system must employ (13)'s variation

$$\tilde{o}\left(n_2 : n_1|r'\right) = \tilde{o}(n_2 : n_1|r)\, \Lambda_{\tilde{o}}\left(r' : r|n \in [n_1, n_2)\right) \in \mathbb{R}$$

$$\forall \left\{n_1, n, n_2, r, r'\right\} \in \mathbb{N}|(1 \leq n_1 \leq n \leq n_2)$$

to recode locally. When it deals with previously encoded data, the local Bayes factor is the classical likelihood factor in statistics:

$$\Lambda_{\tilde{o}}\left(r' : r|n \in [n_1, n_2)\right) = \frac{\ell(r|n)}{\ell(r'|n)} = \frac{\log_n r}{\log_n r'} = \frac{\ln r}{\ln r'}$$

Thus, the degree to which the outcome $n \in [n_1, n_2)$ of the random variable Y supports "hypothesis" r' against r is independent of n. Then, a coding source can change to radix r' utilizing

$$\tilde{o}\left(n_2 : n_1|r'\right) = \tilde{o}(n_2 : n_1|r) \frac{\ln r}{\ln r'} \tag{15}$$

which coincides with the odds of n_2 against n_1 with radix r' for the first time

$$\tilde{o}(n_2 : n_1|r) \frac{\ln r}{\ln r'} = \left(\frac{n_1}{n_2} \log_r \frac{n_2}{n_1}\right) \frac{\ln r}{\ln r'} = \frac{n_1}{n_2} \log_{r'} \frac{n_2}{n_1}$$

Note that the transformation $r \to r'$ increases the odds if $r > r'$, and vice versa, decreases the odds if $r < r'$.

For example, a coding source locally encodes the bin $[4, 13)$ using radix $r = 100$ as

$$\tilde{o}(13 : 4|100) = \frac{4}{13} \log_{100} \frac{13}{4} \approx 0.07875$$

This value measures the entropic contribution of bin $[4, 13)$ to the 100-ary logarithmic scale. When the coding source changes the radix to $r' = 110$, it delivers using (15)

$$\tilde{o}(13 : 4|110) = \tilde{o}(13 : 4|100) \frac{\ln 100}{\ln 110} \approx 0.07715$$

Then, changing the radix to $r' = 90$ yields

$$\tilde{o}(13 : 4|90) = \tilde{o}(13 : 4|110) \frac{\ln 110}{\ln 90} \approx 0.0806$$

Finally, the coding source decodes the odds with radix 90 by solving the prior from (14), i.e.,

$$\tilde{o}(13 : 4) = \frac{\tilde{o}(13 : 4|90)}{\Pr(90, [4, 13))} = \frac{0.0806}{\log_{90} \frac{13}{4}} \approx 0.30769 \approx \frac{4}{13}$$

Remember that local Bayesian coding copes not only with ratios of digits but with ratios of numerals in general. For example, the rational $^{95}/_{971}$ (bin $[95, 971)$) encoded with radix 4 is

$$\tilde{o}(971:95|4) = \frac{1133_4}{33023_4}\log_4\frac{33023_4}{1133_4} \approx 0.02213333_4$$

If environmental conditions cast a change to radix 3, the coding source would decode the datum

$$\tilde{o}(971:95|3) = 0.02213333_4\frac{\ln 4}{\ln 3} \approx 0.01212022_3$$

as

$$\tilde{o}(971:95) = \frac{\tilde{o}(971:95|3)}{\ell([95,971)|3)} \approx \frac{0.01212022_3}{\log_3\frac{1022222_3}{10112_3}} \approx \frac{95}{971}$$

3.3. Elemental Jumps

Elemental jumps between successive quanta or digits are a problem where the synergy between NBL and the odds form of Bayes' rule manifests. Physically, they explain why minor energy levels of a system are more stable than high-energy ones.

The odds (11) between consecutive quanta

$$\tilde{O}(q+1:q|b) = \left(\frac{q}{q+1}\right)^2 \Pr(b,[q,q+1)) = \left(\frac{q}{q+1}\right)^2 \frac{1}{qH_{b-1}} = \frac{q}{(q+1)^2 H_{b-1}}$$

measure the associated harmonic likelihood gap in a given base b, where we have used Equations (3) and (4). b-normalized quantum jumps define the PMF

$$\c{C}_b = \frac{1}{H_{b-1} - H_{b-1,2}}$$

$$\Pr_{\tilde{O}}(b,q) = \c{C}_b\frac{q}{(q+1)^2} \in \mathbb{Q} \tag{16}$$

an exact and multiplicatively separable function where $H_{N,2} \equiv \sum_{k=1}^{N} {}^{1}/_{k^2}$ is the generalized Nth harmonic number in power two and $1 \leq q < b-1$. It is well defined because of

$$\sum_{q=1}^{b-2}\Pr_{\tilde{O}}(b,q) = 1$$

so we can take it as the odds version of PMF (4).

Note that the summation only goes until the penultimate quantum $q = b-2$ because $q = b-1$ cannot jump to b. For example, $\c{C}_4 = {}^{36}/_{17}$, $\Pr_{\tilde{O}}(4,1) = \c{C}_4{}^{1}/_{4} = {}^{9}/_{17}$, and $\Pr_{\tilde{O}}(4,2) = \c{C}_4{}^{2}/_{9} = {}^{8}/_{17}$. With $b = 7$, we obtain $\Pr_{\tilde{O}}(7,1) = {}^{900}/_{3451} \approx 0.261$, $\Pr_{\tilde{O}}(7,2) = {}^{800}/_{3451} \approx 0.232$, $\Pr_{\tilde{O}}(7,3) = {}^{675}/_{3451} \approx 0.196$, $\Pr_{\tilde{O}}(7,4) = {}^{576}/_{3451} \approx 0.167$, and $\Pr_{\tilde{O}}(7,5) = {}^{500}/_{3451} \approx 0.145$. Figure 4 outlines in red the PMF corresponding to standard undecimal in a global setting. The information gap decays harmonically from the second quantum so that transiting from the greatest quanta is easier than from the minor ones. Indeed, only the first few quanta remain stable.

Developing similar reasoning in a local setting, using Equations (6), (7), and (14) ($d_1 = d$ and $d_2 = d+1$), the odds between consecutive digits

$$\tilde{o}(d+1:d|r) = \frac{d}{d+1}\log_r\left(1+\frac{1}{d}\right) \tag{17}$$

measure the associated likelihood gap in radix r. Then, we can calculate the PMF that normalizes these digit gaps in a given radix; the larger the digit, the lesser the information differential. For example, the PMF corresponding to standard quaternary is $\{0.561814, 0.438186\}$. With radix $r = 7$, we obtain $\Pr_{\tilde{o}}(7,1) \approx 29.8\%$, $\Pr_{\tilde{o}}(7,2) \approx 23.24\%$, $\Pr_{\tilde{o}}(7,3) \approx 18.55\%$, $\Pr_{\tilde{o}}(7,4) \approx 15.35\%$, and $\Pr_{\tilde{o}}(7,5) \approx 13.06\%$.

Figure 4 outlines in green the logarithmic PMF of standard undecimal, measuring the improbability of a random local jump through its contribution to the coding source's entropy. The lowest digits maintain discernibility from the environment, while the decreasing entropic support of the more significant digits makes them more vulnerable.

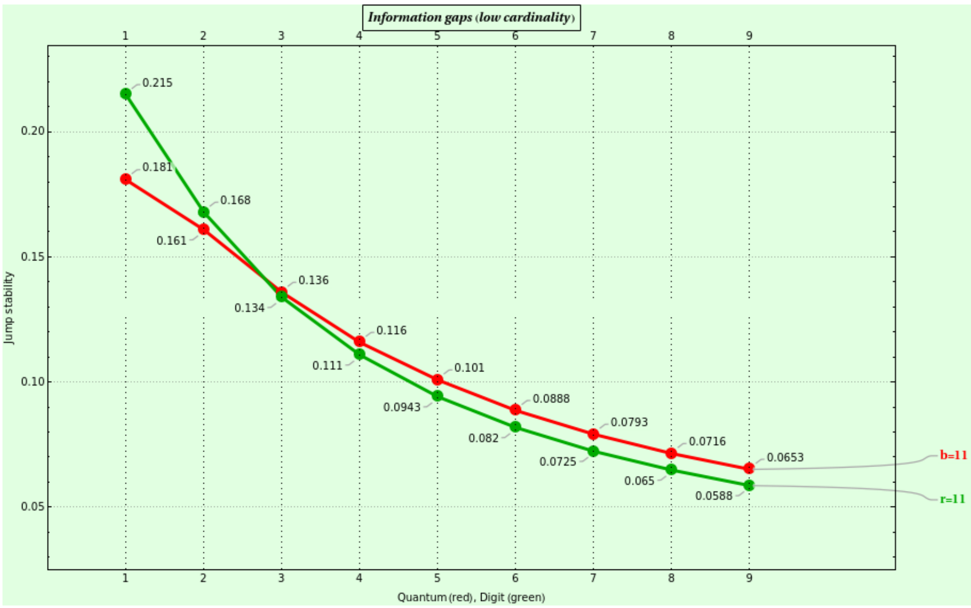

Figure 4. These are the information gaps the undecimal numeral system induces. b stands for standard base, and r for standard radix. Note that the fiducial NBL for decimal numerals (Equation (7)) is steeper than the digit plot (in green, Equation (17)), and the digit plot is steeper than the quantum (in red, Equation (16)).

Although the fiducial NBL is steeper than the corresponding $\Pr_{\tilde{o}}$ regardless of the radix, and this is steeper than $\Pr_{\tilde{O}}$ irrespective of the base, these three plots are hardly distinguishable for large cardinalities (see Figure 5), meaning that an NBL probability is synonymous with stability. A transition from the greatest quanta or digits is much more frequent than a transition from the minor ones. This condition resembles the reactivity of the chemical element periodic table concerning the electron shell (i.e., principal quantum number). More generally, ascending order (of numbers, quanta, digits, or shells) correlates with unsteadiness, which explains why closeness prevails over farness.

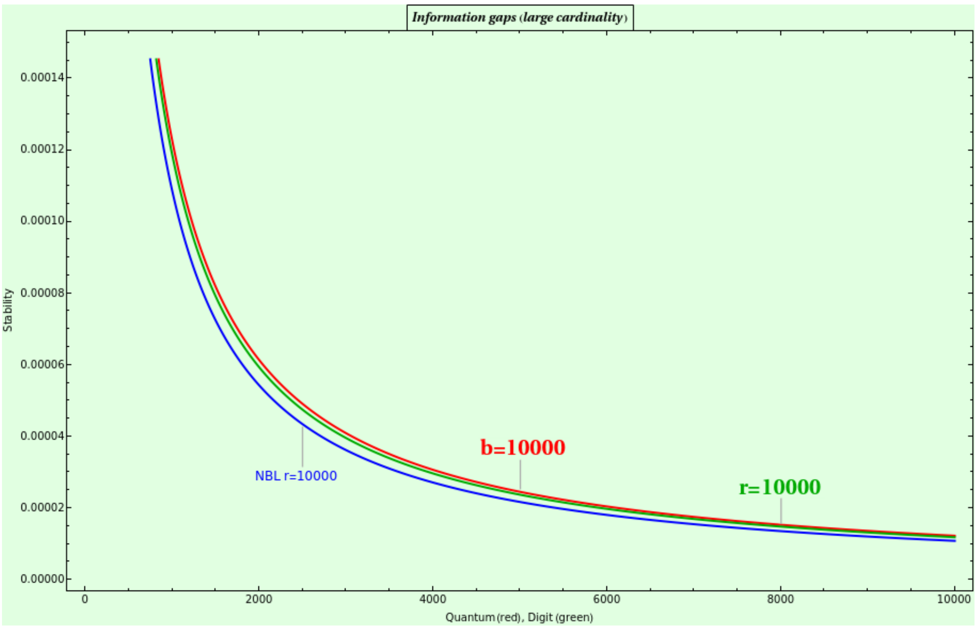

Figure 5. These are information gaps induced by the quanta of global standard base 10,000 (in red, Equation (16)) and the digits of local standard radix 10,000 (in green, Equation (17)), compared with the fiducial NBL (Equation (7)).

3.4. Optimal Stopping

We introduce a class of problems where the synergy between NBL and the odds form of Bayes' rule elucidates a solution.

A PN system assigns an information value to the concepts of likelihood, probability, and odds. In Sections 3.1 and 3.2, we argued that Bayes' rule is the entropic contribution of a bucket to a harmonic scale or a bin to a logarithmic scale. In particular, Equation (14) allows us to calculate the information we can extract from a bipartition by anchoring the first and last domain digits. Assuming $1 \ll r < b$, the local odds of obtaining digit x against 1 and r against x estimate the information aggregate of the two parts. Inherent to X's dichotomy $\{[1,x),[x,r)\}$,

$$\begin{aligned} \ddot{o}_r(x) &= \ddot{o}(\{[1,x),[x,r)\}) \\ &= \tilde{o}(x:1|r) + \tilde{o}(r:x|r) \\ &= \tfrac{1}{x}\log_r x + \tfrac{x}{r}\log_r \tfrac{r}{x} \end{aligned}$$

gives the bipartite odds in logarithmic r-ary units of information, where $1 < x < r$.

We obtain additive countability by making $\kappa \int_1^r \ddot{o}_r(x)\,\mathrm{d}x = 1$. The entropy (local likelihood) distribution function

$$\kappa = \frac{4}{r^2 + 4(r\ln r - 1)\ln\sqrt{r} - 1}$$

$$\overleftarrow{o}_r(x) = \kappa \frac{r}{x} \ln x$$

$$\overrightarrow{o}_r(x) = \kappa x \ln \frac{r}{x}$$

$$\overleftrightarrow{o}_r(x) = \overleftarrow{o}_r(x) + \overrightarrow{o}_r(x) \tag{18}$$

gives the normalized bipartite odds so that $\overleftrightarrow{o}_r(x) = \kappa\ddot{o}_r(x)$ acquires a value between 0 and 1.

For $\overleftrightarrow{o}_r(1) = \overleftrightarrow{o}_r(r) = \kappa \ln r$, both $\overleftrightarrow{o}_r(1)$ and $\overleftrightarrow{o}_r(r)$ tend to vanish in the limit $r \to b \to \infty$. Where does (18) become stationary? When $r \geq 55$, the normalized bipartite odds produce two maxima corresponding to $\overleftarrow{o}_r(x)$ and $\overrightarrow{o}_r(x)$. These maxima optimize the total information transmission of the system. We find at $x = \sqrt{r}$, between the two maxima, a digit that minimizes the distinguishability between the two partitions, which is the analog of the middle point of a segment on the linear scale.

For example, with radix $r = 10,000$, the bipartitions are maximally entropic about $\{[1, 2.7329), [2.7329, 10000)\}$ and $\{[1, 3659.1), [3659.1, 10000)\}$, and $\overleftrightarrow{o}_r(100)$ is the minimum. Figure 6 repeats this exercise and shows the results with $r = 100$ applicants.

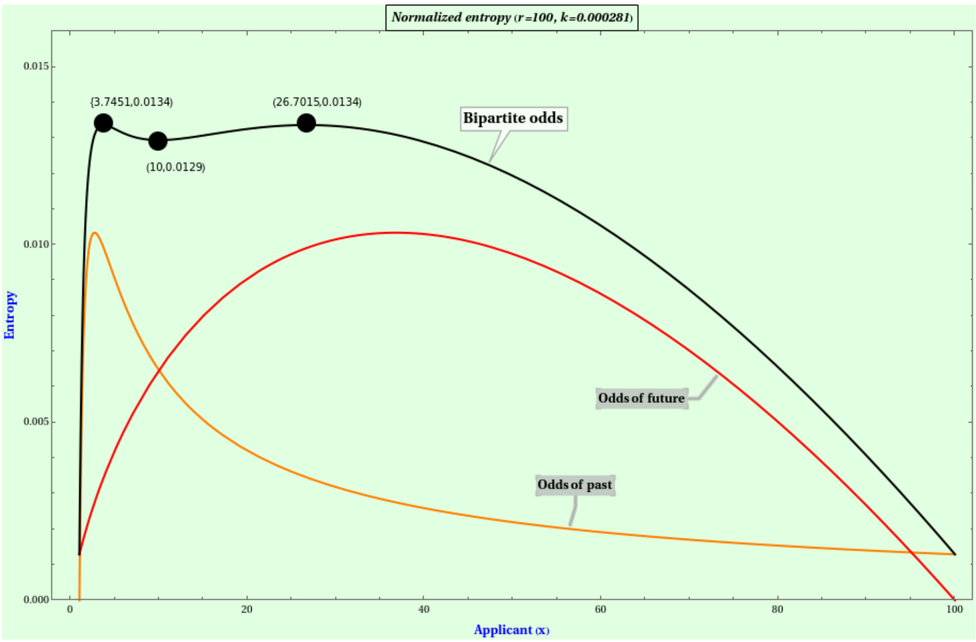

Figure 6. Plots of the odds that yield the past ($\overleftarrow{o}_r(x)$, in yellow), future ($\overrightarrow{o}_r(x)$,in red), and bipartite entropy (the local likelihood distribution function given by (18), in black) with radix 100 (99 applicants). The three points correspond to the maxima and the minimum, whose abscissas give place to the bipartitions with the most information and the lowest bipartite distinguishability, respectively.

Supposing that $1/x$ and x/r are probabilities, Equation (18) is the addition of the corresponding entropies. Both maxima separate a stage of "retention" from a "decision" stage. Retention implies input processing, which raises entropy, whereas decision involves output processing, which lowers entropy. Maximum entropy indicates the most efficient share between the ascent and descent sections. Overall, the plot of $\overleftrightarrow{o}_r(x)$ reflects a natural entropic imbalance toward the small values; $\overleftarrow{o}_r(x)$ dominates in the short term, whereas $\overrightarrow{o}_r(x)$ dominates in the middle and long terms. Computationally, it induces the bulk of processing far before reaching $x = r/2$, while physically, it implies a bias of space or time.

The bipartite odds function can have interesting consequences in computational physics, especially in sequential decision making to solve optimal-stopping (or planning) problems with solutions such as the odds algorithm [52]. Specifically, the secretary problem [53] is a mathematical trope to grasp how computation closely ties with incremental (Bayesian) inference, hence with the asymmetric management of fundamental resources. Shortly, one of $r - 1$ sequentially interviewed applicants must be nominated, with the proviso that they will be either chosen or rejected just after being examined; past the first $\lfloor x \rfloor$ applicants (typically a secretary, but also a lead actor or actress or a car), the judges select the next one that is better than any of the previous ones. $x \approx r/e$ maximizes $\overrightarrow{o}_r(x)$, i.e., the probability of success in choosing "the best" applicant.

Instead, $\overleftrightarrow{o}_r(x)$ answers a different question. What is the optimal size $\lceil x \rceil$ of examined applicants to maximize the odds of choosing "a good" one? This nuance implies a crucial difference in approaching a solution; in this case, we must consider both terms of (18). We define "a good" prospect as a "candidate" in terms of the classic secretary problem, i.e., a seeker (or contender, or claimant) better than the previously examined applicants.

Considering that x is the current applicant, the bipartition separates the past from the future because $\overleftarrow{o}_r(x)$ and $\overrightarrow{o}_r(x)$ focus on the expected benefit before and after x and the practicality of preceding against succeeding data.

Regarding the "past" term, $\ln x$ is the amount of information ascribed to examined applicants, and $1/x$ is the probability of using such information. As $x \to 1$, we take advantage of less and less gathered information, albeit more likely, whereas if $x \to r$, we can leverage more and more references, albeit less likely. There is a compromise between choosing the first applicant (i.e., utterly uninformed decision making) and selecting the last applicant (i.e., ignoring the acquired information). We obtain the maximum of $\overleftarrow{o}_r(x)$ at $x \to \mathrm{e}$ as $r \to b \to \infty$.

Regarding the "future" term, $\ln r/x$ is the information we can obtain from forthcoming applicants, and x/r is the probability of using such information. As $x \to 1$, we will surely miss the most suitable prospects; if $x \to r$, we will hardly find a suitable applicant. There is a compromise between choosing the first applicant (i.e., ignoring the information the remaining applicants can provide) and selecting the last applicant (i.e., remaining information exhausted). We obtain the maximum of $\overrightarrow{o}_r(x)$ at $x \to r/\mathrm{e}$ as $r \to b \to \infty$.

Summing both terms implies balancing the partition behind against the partition ahead. If x is too low ($x \gtrsim 1$), you have the most information ahead for an acceptable selection, and if x is too high ($x \lesssim r$), you have many references for a good choice. Unfortunately, if x is too low ($x \gtrsim 1$), you have less probability of making an acceptable selection, and if x is too high ($x \lesssim r$), you have probably missed the finest choices. While bipartition $\{[1, r/\mathrm{e}), [r/\mathrm{e}, r)\}$ implies a probability of $1/\mathrm{e}$% of skipping and selecting the best alternative, bipartition $\{[1, \mathrm{e}), [\mathrm{e}, r)\}$ reduces this percentage significantly to anticipate a nomination. Thus, $\overleftarrow{o}_r(x)$ enables promptness and $\overrightarrow{o}_r(x)$ quality.

The entropy distribution function of a bipartiton rises to the first maximum, falls and rises again to reach the second maximum, and decays until it almost vanishes. The right holistic strategy is to wait for the information to stop rising so that $\partial \overleftrightarrow{o}_r(x)/\partial x$ vanishes and $\partial^2 \overleftrightarrow{o}_r(x)/\partial x^2$ decreases, i.e., in agreement with the maximum entropy principle for isolated systems (and the minimum energy principle for closed systems) in thermodynamics.

Exclusively concentrating on the past term also makes sense. The idea is to assess the general level after examining only a few applicants. Assuming that ours behaves as a linear time-invariant system, deviations decay as e^{-x}, so the probability that the mean of the three first interviewed applicants is close to the pool mean is $1 - \mathrm{e}^{-3} \approx 95\%$. Since a threesome reasonably represents the whole set of applicants, we can confidently pick a forthcoming candidate. $\overleftarrow{o}_r(x)$ considers the cost of the processing; it is a precursor of human intuition and opens the door to computational methods of solution refinement. For instance, assuming that we can retain a (preliminary) solution, we can progressively renew candidates between the two maxima. If the selection process continues after the second maximum, we are in the same scenario as the classic secretary problem.

Deciding near $x = \sqrt{r}$, between the maxima, is questionable because having already spent substantial resources on obtaining information, the probability of picking the best applicant still needs to reach the optimum. Nonetheless, it is a separator of the two partitions that a living being, for instance, can seek on purpose to maximize internal order or coherence.

The local entropy distribution function has applications other than optimal stopping problems. In general, when the input corresponds to raw data of a natural phenomenon, we can assume NBL statistical redundancy and fit data with (4) or (7). In cryptography, this assumption can facilitate frequency analysis against many ciphers. In data compression, there is no need to scan the whole data block to estimate the underlying PMF; we can randomly and iteratively pick trios or quartets of symbols (past maximum) within the block until satisfying a fitting condition. For instance, (n-ary) Huffman is a lossless data compression prefix code that can presuppose that the global NBL is underneath when the symbol frequencies are unknown. Likewise, "arithmetic coding" could use Benford's cumulative distribution function. If the input is not NBL-compliant, we can scan over a group of approximately r/e symbols (future maximum) to calculate a predetermined probability model. Moreover, adaptive entropy encoders can estimate the probability distribution for the following three or four symbols based on the observed frequencies of the last r/e values.

Let us focus on Golomb coding [54], a lossless data compression method appropriate to alphabets following a geometric distribution. It pivots on an adaptative parameter M to split an input value into a twofold length-variable sequence of bits, to wit, the quotient q of a division by M, sent in unary coding, and the remainder ξ, sent in truncated binary encoding. We must fix M to an integer value other than 1 (no compression). For example, suppose we want to encode the decimal $N = 67$ with $M = 10$ (Rice code bits $b = \lfloor \log_2(M) \rfloor = 3$). Then, $q = \lfloor N/M \rfloor = \lfloor 67/10 \rfloor = 6$ and $\xi = 7$, delivering

$$\begin{aligned} quotientCode(q), remainderCode(r) &= code_q(\lfloor N/M \rfloor), code_\xi(N \text{ modulo } M) \\ &= code_q(6), code_\xi(7) \\ &= 1111110, 1101 \end{aligned}$$

where 1111110 is a q-length string of ones with suffix 0 and 1101 is the binary prefix code representation of ξ using $b + 1$ bits.

We can modify this algorithm to establish a double Benford–Bayes parameter. $M_{past} = 4$ ($b_{past} = \lfloor \log_2(M_{past}) \rfloor = 2$). Then, $q_{past} = \lfloor N/M_{past} \rfloor = \lfloor 67/4 \rfloor = 16$ and remainder $\xi_{past} = 3$, delivered as 11. Then, we further divide the past quotient code into another quotient–remainder pair fixing $M_{future} = \lceil q_{past}/e \rceil = 6$ ($b_{future} = \left\lfloor \log_2\left(M_{future}\right) \right\rfloor = 2$), producing $q_{future} = \lfloor q_{past}/M_{future} \rfloor = 2$ (delivered as 110) and $\xi_{future} = 4$ (delivered as 100), so the final code would turn out to be 110, 100, 11. Both the encoder and the decoder assume this Benford–Bayes "model", so that we do not have to spend resources to determine the optimal value of M when the probability distribution for integers is unknown, nor do streaming data require a first pass over the data to define a probability model. Moreover, since we have a fixed model, the encoder and decoder do not need synchronization or waste computational power to adapt their state.

3.5. Gap Ratio and Cross-Ratio

If we call the quotient

$$\frac{A-B}{C-D} \equiv (A : B \,|\, C : D)$$

a "ratio of gaps" or "gap ratio", we can rewrite the encoding rule (11) as

$$\tilde{O}(t : s|b) = \left(s^2 H_{t-1} : s^2 H_{s-1} \,|\, t^2 H_{b-1} :\right) \in \mathbb{Q} \; (1 \leq s < t < b)$$

Interestingly, the product, quotient, sum, and difference of two gap ratios are gap ratios,

$$(A : B \,|\, C : D) + (J : K \,|\, L : M) = \frac{A-B}{C-D} + \frac{J-K}{L-M} = \tag{19}$$

$$= \frac{(AL + BM + CJ + DK) - (AM + BL + CK + DJ)}{(CL + DM) - (CM + DL)}$$

$$(A : B \,|\, C : D) - (J : K \,|\, L : M) = \frac{A - B}{C - D} - \frac{J - K}{L - M} = \tag{20}$$

$$= \frac{(AL + BM + CK + DJ) - (AM + BL + CJ + DK)}{(CL + DM) - (CM + DL)}$$

$$(A : B \,|\, C : D) \times (J : K \,|\, L : M) = \frac{A - B}{C - D} \times \frac{J - K}{L - M} = \tag{21}$$

$$= \frac{(AJ + BK) - (AK + BJ)}{(CL + DM) - (CM + DL)}$$

$$(A : B \,|\, C : D) \div (J : K \,|\, L : M) = \frac{A - B}{C - D} \div \frac{J - K}{L - M} = \tag{22}$$

$$= \frac{(AL + BM) - (AM + BL)}{(CJ + DK) - (CK + DJ)}$$

In other words, the set of gap ratios, $\widetilde{\mathbb{Q}}$, is an ordered algebraic field (of characteristic zero). We can represent a gap ratio $(A : B \,|\, C : D)$ on a square grid as an oriented rectangle with vertices (A, D), (B, D), (A, C), and (B, C) (see Figure 7).

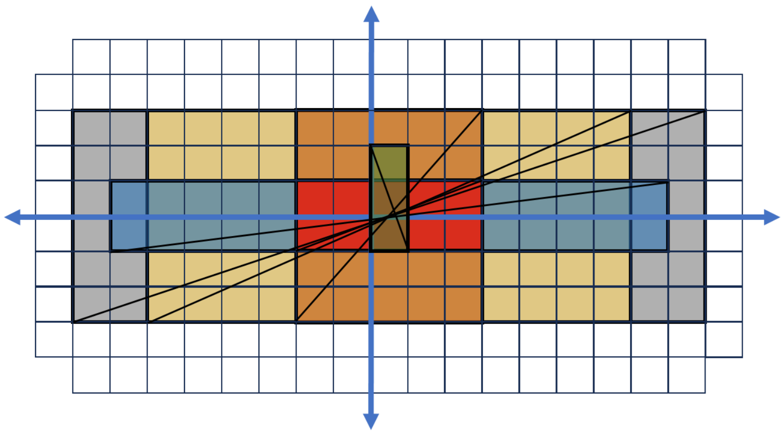

Figure 7. Example of representing a gap ratio and its arithmetic. We display the gap ratios $(-2 : 3 \,|\, 1 : -1)$ (in red) and $(1 : \,|\, 2 : -1)$ (in green) and the result of the basic operations between them, i.e., addition $(-6 : 7 \,|\, 3 : -3)$ in yellow (19), subtraction $(-8 : 9 \,|\, 3 : -3)$ in gray (20), multiplication $(-2 : 3 \,|\, 3 : -3)$ in orange (21), and division $(-7 : 8 \,|\, 1 : -1)$ in blue (22). The diagonals from bottom left to top right indicate a negative sign, while the diagonal from top left to bottom right indicates a positive sign.

We can formally define gap ratios as equivalence classes (symbol "$\sim$") of rational quadruplets where $(A : B \,|\, C : D) \sim (K : J \,|\, L : M) \iff (A - B)(L - M) = (C - D)(J - K)$. Note that $\mathbb{Q} \subset \widetilde{\mathbb{Q}}$ due to $N/D \equiv (N : \,|\, D :)$.

This view lets us comprehend a gap ratio as a rational number with length or area. So, $\widetilde{\mathbb{Q}}$ is an extension field of $\mathbb{Q}$ such that the latter's operations are a particular case of the former. Specifically, the multiplicative units of the field are $(1 : \,|\, 1 :)$ and $(: 1 \,|\, 1 :)$. The reciprocal of $(J : K \,|\, L : M)$ is $(L : M \,|\, J : K)$. In addition, $(J : K \,|\, L : M) + (K : J \,|\, L : M)$ and $(J : K \,|\, L : M) - (K : J \,|\, M : L)$ vanish.

Local Bayesian data also have the structure of a gap ratio. The original prior odds are precisely the rational n_1/n_2. We can rewrite the encoding rule (14) as

$$\tilde{o}(n_2 : n_1|r) = (n_1 \ln n_2 : n_1 \ln n_1 \,|\, n_2 \ln r :) \in \mathbb{R}\ (1 \leq n_1 < n_2)$$

Likewise, the result of comparing a pair of local odds is the gap ratio

$$\tilde{o}(j : i|h : g) = \frac{\tilde{o}(j : i|r)}{\tilde{o}(h : g|r)} = \frac{\frac{i}{j}\log_r \frac{j}{i}}{\frac{g}{h}\log_r \frac{h}{g}} = (hi \ln j : hi \ln i \,|\, gj \ln h : gj \ln g) \in \mathbb{R}$$

where $(1 \leq g < h) \wedge (1 \leq i < j)$.

Bayesian data are elements of $\widetilde{\mathbb{Q}}$. Certain elements within this set are of special interest; the following double gap ratio is the "cross-ratio" of four distinct points [55]:

$$\frac{(A : C \,|\, A : D)}{(B : C \,|\, B : D)} = \frac{\frac{A-C}{A-D}}{\frac{B-C}{B-D}} \equiv (A, B; C, D) \tag{23}$$

where the alphabetical order indicates that A, B, C, and D are consecutive on the rational projective line, and $A - B$ and $C - D$ have the same sign. The cross-ratio characterizes the projective line's geometry, calculating how much the quadruple's crossing symmetries deviate from the ideal proportion 1, precisely the extent to which the ratio of how C divides $[A, B)$ is proportional to how D divides $[A, B)$.

We can simplify the notation when one of the cross-ratio points is infinite:

$$(A, \infty; C, D) = (A : C \,|\, A : D) \equiv (A; C : D) = \frac{A - C}{A - D}$$

Moreover, we can write

$$(A, B; C, D) = (AB + CD; AD + BC : AC + BD)$$

using (22) and (23).

In physics, we must generically understand the concept of correlation as a ratio between magnitudes of the same physical unit. The most straightforward embodiment of the gap ratio and the cross-ratio gives the Doppler effect's relationship between the frequency perceived by the receiver f_r and the emitted frequency f_s, i.e.,

$$\frac{f_r}{f_s} = \frac{s_w - s_r}{s_w + s_s} = (s_w; s_r : -s_s)$$

where s_w is the propagation speed of waves in the medium, $s_r < s_w$ is the speed of the receiver relative to the medium, and $s_s < s_w$ is the speed of the source relative to the medium, assuming that they are moving away from each other [56]. Likewise, the formula of the relativistic Doppler effect of the source's frequency relative to the receiver's frequency moving away at speed v is [57]

$$\left(\frac{f_s}{f_r}\right)^2 = \frac{1 + v/c}{1 - v/c} = (1; -v/c : v/c) \tag{24}$$

4. Conformality

Departing from an inverse-square PMF for the naturals, we gleaned the global and local NBL, implying that a double scale is necessary to support a universal place-value system. A global base specifies the harmonic scale, while a local radix fixes the logarithmic scale that a coding source uses to represent numerals in PN.

Bayesian data are elements of the field of gap ratios. The logarithm of a cross-ratio, an especially useful subclass of gap ratio, determines the conformal metric of local coding spaces.

Conformal maps preserve angles, hence the shapes of the figures, which also implies scale invariance. These properties are critical to translating the elements of a global harmonic space into a local logarithmic subspace, the latter reflecting the state of the former. Conformality is a requirement for coding information that drives complexity; "there is a shared very particular characteristic of all complex systems. And that is they internally encode the world in which they live" [58].

4.1. The Conformal 1-Ball Model

The cross-ratio paves the way to conformality because it is invariant under linear fractional transformations over rings [59]. The group of linear fractional transformations $^{(Az+B)}/_{(Cz+D)}$, where $\{A, B, C, D\} \in \mathbb{Z}$, called the "modular group" (a subset of the Möbius group), acts transitively on the points of the grid $\mathbb{Z}^2$ visible from the origin, i.e., the irreducible fractions [60], so preserving the form of polygonal shapes through the cross-ratio.

Because the harmonic and logarithmic scales handle the concept of cross-ratio, we can find a modular transformation between four specific points in a global space S and four points in a given S's subspace, the coding space where the source makes a local model.

The most powerful application of the cross-ratio is the Poincaré disk (The Non-Euclidean World in [61]), a conformal model of hyperbolic geometry that projects the whole H^2 in the unit disk. Circle-preserving Möbius transformations are the isometries of the complex plane. Assuming that the disk center is at the plane's origin, points z_2 and z_3 within the disk connected by the arc of a geodesic circle perpendicularly intersecting the disk's boundary at z_1 and z_4 are at a hyperbolic distance of $\ln(z_1, z_2; z_3, z_4)$ [62]. This measure is invariant under the subset of Möbius maps acting transitively on the unit disk, the space of the coding source.

In one dimension, the complex plane augmented by the point at infinity can be considered the real projective line [63], and the disk becomes the unit 1-ball. More specifically, the set of irreducible fractions augmented by the point at infinity is the rational projective line; hence, the unit 1-ball becomes the rational open unit interval.

While $(-1, 1)$ is the mathematical domain where the modular group acts, we are interested in the global computational space where Bayesian processes and transformative calculation methods occur. We assume that global Bayesian data, i.e., rational quanta, populate a cosmos of information a source perceives and codes to create a continuous world model. Outside a coding source, the information resides on a harmonic scale, whereas inside, a logarithmic scale lodges local Bayesian data.

Suppose that an object is at position P outside $(-1, 1)$. We are ignorant of the actual computation of P, but we know that it is a rational resulting from applying the rule (11). Be that as it may, we can use Equation (14) to locally figure the odds of $P-1$ against $P+1$ in radix r, whose Bayes factor is the logarithm in r-ary units of a cross-ratio where $z_1 = -1$, $z_2 = 1$, $z_3 = P$, and $z_4 = \infty$, i.e.,

$$\log_r(-1, 1; P, \infty) \equiv \log_r(P; -1 : 1) = \log_r \frac{P+1}{P-1}$$

In information theory, this expression is the representational length in radix r of the rational number $^{P+1}/_{P-1}$ and, according to NBL (8), the r-normalized width of bin $[P-1, P+1)$. We can unite these outlooks by interpreting this Bayes factor as the hyperbolic distance from P to $b \to \infty$, i.e.,

$$d_r(P, \infty) \equiv \log_r(-1, 1; P, \infty) = \frac{2}{\ln r} \operatorname{arcoth}(P)$$

where arcoth is the inverse function of the hyperbolic cotangent.

A neat inversion conformally maps the outside of the coding source to its inside,

$$(-1,1;P,\infty) \underset{z\mapsto 1/z}{\longrightarrow} (-1,1;D,0) \tag{25}$$

conserving the cross-ratio. (Other inversions $z \mapsto\propto 1/z$ also serve but violate the minimal information principle.) For example, if $P = 2$ (hence, $D = 1/2$), $(-1,1;P,\infty) = \frac{-1-2}{1-2} = 3 = \frac{1+\frac{1}{2}}{1-\frac{1}{2}} = (-1,1;D,0)$. Therefore, the r-normalized hyperbolic distance between the origin and $|D| = |1/P| < 1$ is

$$d_r(0,D) \equiv \log_r(-1,1;D,0) \equiv \log_r(0,D;1,-1) \equiv$$

$$\equiv \log_r \frac{1+D}{1-D} \equiv \log_r(1;-D:D) \equiv \frac{2}{\ln r}\mathrm{artanh}(D)$$

where artanh is the inverse function of the hyperbolic tangent, and $\ln r$ is the (constant) curvature's absolute value of the source's coding space.

The inverse function of $d_r(0,1/P)$ is $P = \coth\left(\left(\frac{1}{2}\ln r\right)d_r(0,1/P)\right)$. For example, an object at a Euclidean distance $P = 10^6$ from the origin is at a natural hyperbolic distance of $d_e(P,\infty) = 2\,\mathrm{arcoth}\left(10^6\right) \approx 2\times 10^{-6}$ from $b \to \infty$. The coding source positions the object at $D = 1/P = 10^{-6}$, at a natural hyperbolic distance of $d_e(0,D) = 2\,\mathrm{artanh}\left(10^{-6}\right) = 2\,\mathrm{arcoth}\left(10^6\right)$ from the origin, and decodes it as $P = \coth\mathrm{artanh}\left(10^{-6}\right) = 10^6$.

Mind that the local coding space is the 1-annulus

$$å^- \equiv \{D \in \mathbb{Q}| -1 < D < -1/b\}$$

$$å^+ \equiv \{D \in \mathbb{Q}|\, 1/b < D < 1\}$$

$$å \equiv å^- \cup å^+ \equiv \{D \in \mathbb{Q}|\, 1/b < |D| < 1\}$$

reflecting what the source observes in the 1-annulus

$$Å^- \equiv \{P \in \mathbb{Q}| -b < P \leq -1\}$$

$$Å^+ \equiv \{P \in \mathbb{Q}|\, 1 \leq P < b\}$$

$$Å \equiv Å^- \cup Å^+ \equiv \{P \in \mathbb{Q}|\, 1 \leq |P| < b\} \tag{26}$$

For instance, if $D \in å^+$, $d_r(0,D)$ vanishes if $P \to b \to \infty$, is > 1 if P is at a Euclidean distance closer than coth $\ln r/2$ from the origin, and diverges if $P \to 1^+$.

The r-normalized hyperbolic distance between two points A and B in å is $d_r(0,Q_B) - d_r(0,Q_A)$, i.e.,

$$d_r(Q_A,Q_B) = \frac{2}{\ln r}(\mathrm{artanh}(Q_B) - \mathrm{artanh}(Q_A)) \tag{27}$$

where Q_A and Q_B result from the conformal transformation (25), i.e., $P \mapsto 1/Q$, which mirrors the external world concerning å's outward boundary, to wit ± 1. Thus,

$$d_r(Q) = d_r(0,Q) = \frac{2}{\ln r}\mathrm{artanh}(Q) \tag{28}$$

reflects how far an object at Q is from infinity, situated at the origin.

Nonetheless, we want the origins of the coding source and Å to coincide and ± 1 to be the infinite points of the local model. This requirement implies calculating Q's complement to one, a logical negation that varies on the left and the right. Recall that all negations are derivations of the canonical one [64], so we will use the map $z \mapsto -1 - 1/z$ on the left and $z \mapsto 1 - 1/z$ on the right to satisfy the minimal information principle. The coding space is now the open 1-ball

$$ß \equiv \{Q \in \mathbb{Q} | \, |Q| < 1 - 1/b\} \tag{29}$$

reflecting what the source observes in the 1-annulus Å.

Equations (27) and (28) are also valid in ß. An object at $P \in Å^-$ will be in $ß^-$ at a Euclidean distance of

$$Q = (;P + 1 : -P) = -1 - \frac{1}{P}$$

from the origin and hence at a hyperbolic distance of

$$d_r^-(Q) = \frac{2}{\ln r}\text{artanh}(Q) = -\log_r(-(2P+1)) = \log_r(P; P+1 : -P-1) \tag{30}$$

with inverse (decoding) function $P = -\frac{1}{2}\left(1 + r^{-d_r^-(Q)}\right)$.

Similarly, an object at $P \in Å^+$ will be in $ß^+$ at a Euclidean distance of

$$Q = (P; 1 :) = 1 - \frac{1}{P}$$

from the origin and hence at a hyperbolic distance of

$$d_r^+(Q) = \frac{2}{\ln r}\text{artanh}(Q) = \log_r(2P - 1) = \log_r(P; 1 - P : P - 1) \tag{31}$$

with inverse (decoding) function $P = \frac{1}{2}\left(1 + r^{d_r^+(Q)}\right)$.

For example, a coding source places an object observed at a Euclidean distance of $P = \pm 10^6$ at $Q = \pm 0.999999$ in ß at a natural hyperbolic distance $d_e(Q) = 2\,\text{artanh}(\pm 0.999999) = \pm \ln\left(\pm 2 \times 10^6 - 1\right) = \pm 14.50866$ from the origin, and $P = \pm\frac{1}{2}\left(1 + e^{14.50866}\right) = \pm 10^6$. On the positive side, the odds are $\tilde{o}(2P - 1 : 1|e) = \frac{\ln(2P-1)}{2P-1} = \frac{\ln 1999999}{1999999} \approx 7.254332 \times 10^{-6}$. Suppose that, later, the coding source calculates this value as 7.254265×10^{-6}, meaning that either the object has moved to $P \approx \pm 1000010$ (because $\tilde{o}(2P - 1 : 1|e) = \frac{\ln 2000019}{2000019} \approx 7.254265 \times 10^{-6}$) or the radix has changed to $r \approx 2.71831$ (because $\tilde{o}\left(2 \times 10^6 - 1 : 1|2.71831\right) = \frac{\ln 1999999}{1999999 \ln 2.718307} \approx 7.254265 \times 10^{-6}$). Even a combination of these two cases could produce the same odds value.

Note that functions in the form $|\theta_1| \exp(|\theta_2 Q|)\text{artanh}^{\theta_3}(|\theta_4|Q)$, where $\{\theta_1, \theta_2, \theta_4 \in \mathbb{Q}\}$ and θ_3 is an odd power, e.g., $2\exp(3Q)\text{artanh}^5(7Q)$, also give rise to an odd hyperbolic distance that complies with boundary conditions $d_r(\pm 1) = \pm\infty$ and a vanishing distance when Q vanishes ($|P| \to 1$). However, they would introduce new factors and parameters we cannot explain; $\text{artanh}(Q)$ is the only conformal function that retains the origin and conforms with the minimal information principle. In addition, it agrees with the canonical PMF (the first power is the most probable) and satisfies the additional condition of having a non-vanishing derivative at the origin, i.e., the origin is not stationary so that the function can keep its increasing tendency from left to right.

4.2. Conformal Coding and Computability

artanh has a protagonist role in a conformal space not only due to its manifestations in physics, mainly the metric (28), but also because its Taylor series allows the natural logarithm itself to be calculated iteratively based on the odd powers of $\frac{x-1}{x+1}$ ([65], 4.1.27), i.e.,

$$\ln x = 2\sum_{i=1}^{\infty} \frac{(x;1:-1)^{2i-1}}{2i-1} \tag{32}$$

This is valid for any $x \in \mathbb{R}^+$, especially when $x \approx 1$. For example, let us calculate the ternary logarithm of $P = 10^6$. Since $10^6 = 1212210202001_3$, a numeral with 13 digits, its logarithm's characteristic is 12 and $x = P - 3^{12} = 1.88168$; after five iterations, the mantissa's error is less than one millionth, i.e., we calculate $\log_3 P = 12 + \frac{2}{\ln 3}\sum_{i=1}^{5} \frac{(1.88168;1:-1)^{2i-1}}{2i-1} = 12.57541925$ against the real value of 12.57541965.

Because the coding source can autonomously calculate the logarithm using (29), ß's curvature $-\ln r$ is a built-in value. Likewise, ß's Euler characteristic $\chi = \int_0^1 \ln x \, dx = -1$, a topological invariant [66] corresponding to no vertices, one edge, and no faces, is a built-in value. Moreover, the coding source is "aware" of the PMF (1) through the digamma function $\psi(x) = \Gamma'(x)/\Gamma(x)$ (see Section 2.3) because the gamma function results from integrating the powers of ß's curvature over the unit segment, namely, $\Gamma(n+1) = \int_0^1 (-\ln x)^n \, dx$.

Let us denominate the r-normalized hyperbolic distance (Equations (31) and (30)) in logarithmic terms the "conformal encoding function" of $P \in$ Å, namely,

$$\overrightarrow{C}_r(P) \equiv \operatorname{sgn}(P)\ \log_r(2\operatorname{sgn}(P)\,P - 1) \tag{33}$$

with inverse "conformal decoding function"

$$\overleftarrow{C}_r(C) \equiv \frac{1}{2}\operatorname{sgn}(C)\left(1 + r^{\operatorname{sgn}(C)\,\overrightarrow{C}_r(P)}\right) \tag{34}$$

where sgn() is the signum function (see Figure 8).

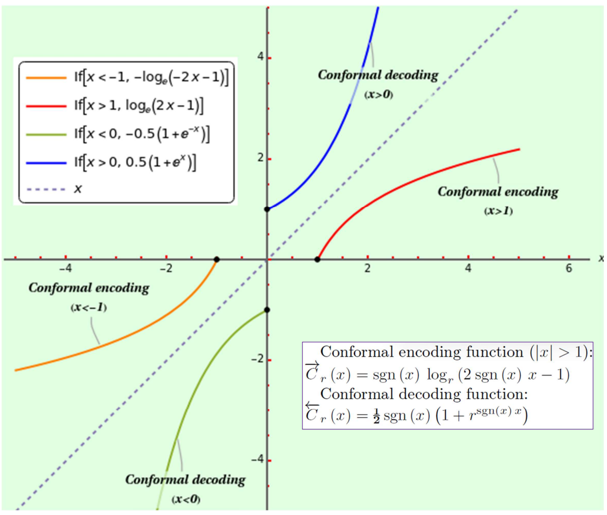

Figure 8. The coding functions (Equations (33) and (34)) of the 1-ball conformal model.

Because the source places an object observed at a Euclidean distance $P \in$ Å at a Euclidean distance

$$Q = \operatorname{sgn}(P) - \frac{1}{P} = \operatorname{sgn}(P)\,(|P|;1:) \tag{35}$$

from the origin, we can calculate the conformal encoding function using (32) as the infinite summation

$$\overrightarrow{C}_r(Q) = \frac{2}{\ln r} \sum_{i=1}^{\infty} \frac{Q^{2i-1}}{2i-1} \tag{36}$$

Note that a coding source can calculate its first digit probability as

$$\Pr(r,1) = \log_r\left(1+\frac{1}{1}\right) = \frac{\int_0^1 \operatorname{artanh}(Q)\,\mathrm{d}Q}{\ln(r)} = \frac{1}{2}\int_0^1 \overrightarrow{C}_r(Q)\,\mathrm{d}Q$$

The coding source can consequently calculate the conformal decoding function as the infinite product

$$\overleftarrow{C}_r(Q) = \frac{1}{2}\operatorname{sgn}(Q)\left(1 + r^{\operatorname{sgn}(Q)\,\overrightarrow{C}_r(Q)}\right) = \frac{1}{2}\operatorname{sgn}(Q)\left(1 + \prod_{i=1}^{\infty} \mathrm{e}^{2\operatorname{sgn}(Q)\frac{Q^{2i-1}}{2i-1}}\right) \tag{37}$$

which does not depend on r.

Therefore, Euclidean distances measured in ß (29) are the only inputs necessary to compute the coding functions.

4.3. Conformal Relativity

We must take the hyperbolic distance (28) as an abstract concept that does not have to be a physical length.

Imagine that the global base b physically represents the speed of light. On the right hand, $P = (b;: v) = {}^{b}/_{(b-v)} \equiv {}^{1}/_{(1-v/c)}$ produces $Q = 1 - {}^{1}/_{P} = {}^{v}/_{b} \equiv {}^{v}/_{c}$; if an object's speed is $v \to b \equiv c$, then $P \to \infty^+$ and $Q \to 1^-$, and if $v \to 0^+$, we obtain $P \to 1^+$ and $Q \to 0^+$. On the left hand, $P = (;b : -(b+v)) = -{}^{b}/_{(b+v)} \equiv -{}^{1}/_{(1+v/c)}$ produces $Q = -1 - {}^{1}/_{P} = {}^{v}/_{b} \equiv {}^{v}/_{c}$; if an object's speed is $v \to -b \equiv -c$, then $P \to \infty^-$ and $Q \to -1^+$, and if $v \to 0^-$, we obtain $P \to -1^-$ and $Q \to 0^-$. In either case, $Q = {}^{v}/_{b}$, and we can write the relativistic Doppler effect (24) in the form

$$\frac{f_s}{f_r} = \mathrm{e}^{\operatorname{artanh}(v/b)} = \mathrm{e}^{\operatorname{artanh}(Q)}$$

Since the rapidity corresponding to velocity v is, by definition,

$$\nu \equiv \operatorname{artanh}(v/b) \tag{38}$$

special relativity's Lorentz factor is $\gamma \equiv \cosh \nu$ [67]. Moreover, the coding source resolves the composition of Doppler shifts as the exponential of the addition of rapidities, i.e.,

$$\frac{f_s}{f_m}\frac{f_m}{f_r} = \mathrm{e}^{\nu_{ms}+\nu_{rm}}$$

where f_m is the frequency perceived by the first receiver, ν_{ms} the rapidity of the first receiver relative to the source, and ν_{rm} the rapidity of the second receiver relative to the first one.

The special relativity theory is only conformal in terms of rapidity. Visualize two inertial frames, A and B, cruising at relativistic speed ratios of $P_A = 3$ and $P_B = 18$ about the origin of the coding source. These correspond in ß (using (35)) at ratios $Q_A = 1 - {}^{1}/_{3} = {}^{2}/_{3}$ and $Q_B = 1 - {}^{1}/_{18} = {}^{17}/_{18}$ of the speed v to $b \equiv c$, defining rapidities $\nu_A = \operatorname{artanh}({}^{2}/_{3})$ and $\nu_B = \operatorname{artanh}({}^{17}/_{18})$, and encoded in ternary as hyperbolic (relativistic) speeds $d_3(Q_A) = {}^{2\nu_A}/_{\ln 3} = 1.465$ and $d_3(Q_B) = {}^{2\nu_B}/_{\ln 3} = 3.2362$. The difference in hyperbolic speeds is linear in ß; using (27), $d_3(Q_A, Q_B) = d_3(Q_B) - d_3(Q_A) = 1.7712$. Within Å

(26), the difference in (Euclidean) velocities is $18 - 3 = 15$, but the difference in hyperbolic speeds is, using (34),

$$\frac{1}{2}\left(1 + 3^{\frac{2}{\ln 3}(\nu_B - \nu_A)}\right) = \frac{1}{2}\left(1 + 3^{1.7712}\right) = 4$$

Rapidity arithmetic is more straightforward than calculating Einstein's subtraction formula of (Euclidean) velocities, which calculates $\nu_B - \nu_A$ as $\operatorname{artanh}\left(\frac{1 - Q_A Q_B}{Q_B - Q_A}\right)$.

Another way to obtain the same result is directly using the cross-ratio and (33), i.e.,

$$\frac{1}{2}\left(1 + 3^{\log_3 \frac{\frac{Q_A - 1}{Q_A + 1}}{\frac{Q_B - 1}{Q_B + 1}}}\right) = \frac{1}{2}\left(1 + 3^{\log_3 \frac{2P_B - 1}{2P_A - 1}}\right) = \frac{1}{2}\left(1 + 3^{\log_3 7}\right) = 4$$

These results mean the weave of Lorentz invariance, and more generally Poincaré invariance, is the algebraic field of cross-ratios. Lorentz symmetry [68] locally preserves central reflections and boosts, the latter maintaining constant the speed of light (the global base) when transforming to a reference frame with a different velocity. Poincaré symmetry, the entire symmetry group of any relativistic field theory, additionally preserves the laws of physics for inertial coding sources situated at different quantum positions.

4.4. Local Bayesian Entropy

The conformal encoding function (33) represents the likelihood of the local Bayesian odds, namely,

$$\tilde{o}(2\,\operatorname{sgn}(P)\,P - 1 : 1\,|\,r) = \frac{\log_r(2\,\operatorname{sgn}(P)\,P - 1)}{2\,\operatorname{sgn}(P)\,P - 1} = \frac{\log_r(2|P| - 1)}{2|P| - 1}$$

which expresses the entropic contribution of bin $[1, 2|P| - 1)$, hence of P, to Å's information total, where (26) defines Å. Because a cross-ratio is invariant under a conformal transformation, so is the Bayesian information defined by the local odds. The transformed Bayesian datum is

$$\tilde{o}(1 + \operatorname{sgn}(Q)\,Q : 1 - \operatorname{sgn}(Q)\,Q\,|\,r) = \frac{1 - \operatorname{sgn}(Q)\,Q}{1 + \operatorname{sgn}(Q)\,Q}\log_r\left(\frac{1 + \operatorname{sgn}(Q)\,Q}{1 - \operatorname{sgn}(Q)\,Q}\right)$$

which expresses the entropic contribution of bin $[1 - |Q|, 1 + |Q|)$, hence of Q, to ß's information total, where (29) defines ß. Moreover, because the mapping is bijective, Å and ß contain the same absolute likelihood information.

The limiting function of the rationals in Å to approximate a piece of "real" average information would require an analysis analogous to [69] (chapter 4*b*), which pivots on the differential entropy [70]. Assuming $b \to \infty$, such a "differential Bayesian entropy" measures the continuous weighted likelihood from the coding source boundary to a point P; it is precisely the integral

$$\tilde{e}_{\text{Å}^+}(P) = \int_1^P \frac{\log_r(2\,x - 1)}{2x - 1}\,\mathrm{d}x = \frac{1}{\ln r}\left(\frac{\ln(2P - 1)}{2}\right)^2$$

on the right and

$$\tilde{e}_{\text{Å}^-}(P) = \int_P^{-1} \frac{\log_r(-2x - 1)}{(-2x - 1)}\,\mathrm{d}x = \frac{1}{\ln r}\left(\frac{\ln(-2P - 1)}{2}\right)^2$$

on the left.

Then, using (35), artanh comes up again to estimate the dominating element of the coding source's entropy

$$\widetilde{e}_{ß}(Q) = \widetilde{e}_{\mathring{A}}(P) = \frac{\operatorname{artanh}^2(Q)}{\ln r}\text{nat}$$

from the origin to infinite points ($Q \to 1$). In the special theory of relativity, this result means that the entropy grows quadratically with the rapidity (38) when $v \to b \equiv c$. Using (28),

$$\widetilde{e}_{ß}(Q) = \ln r \left({}^{d_r(Q)}/2 \right)^2 nat$$

in terms of distance; note that this expression peers the Bekenstein–Hawking's formula of black hole entropy in quantum gravity [71]. Since entropy measures confusion, this result means that objects in remarkably curved coding spaces or at huge distances are indiscernible.

4.5. Conformal Iterated Coding

Because the hyperbolic distance (Equations (28) or (33)) and the rapidity range between $-\infty$ and ∞, we can presume the coding source's logarithmic scale is a (new) whole external world, defined by Euclidean distances, and repeat the encoding process.

If we apply the (right-hand) conformal encoding function recursively,

$$\log_r(2\log_r(2\log_r(2\ldots-1)-1)-1)$$

the source encodes a rightward external object's position sooner or later in $ß^+$, and the recursion halts. For example, if $r = 3$, we can map an object observed at a Euclidean distance of googol from the origin, after five nested conformal transformations, onto a point in $ß^+$ at an approximated hyperbolic distance of 0.096773 from the origin.

Repeatedly applying the encoding function (33) or the decoding function (34) is information-preserving iterated coding. We will use the notation $r \circ n$ to express the nth iterate ($n \geq 1$) of the encoding function $\overrightarrow{C}_r(P)$, so that $\overrightarrow{C}_{r\circ 1}(P) \equiv \overrightarrow{C}_r(P)$ and

$$\overrightarrow{C}_{r\circ(n+1)}(P) \equiv \overrightarrow{C}_r(P) \circ \overrightarrow{C}_{r\circ n}(P)$$

where "$\circ$" denotes function composition holding the properties

$$\overrightarrow{C}_{r\circ(m+n)}(P) \equiv \overrightarrow{C}_{r\circ n}(P) \circ \overrightarrow{C}_{r\circ m}(P)$$

and

$$\overrightarrow{C}_{r\circ mn}(P) \equiv \overrightarrow{C}_{r\circ n}\left(\overrightarrow{C}_{r\circ m}(P)\right)$$

Note that the limits of the coding space remain unaltered irrespective of the iteration because $\overrightarrow{C}_{r\circ n}(\pm 1)$ vanishes for all n.

The iterated logarithm of $N \in \mathbb{N}^+$, written $\log_* N$, is the number of times the natural logarithm function must be recursively applied before the result is less than or equal to the unit. Similarly,

$$\overrightarrow{C}_{r*}(P)$$

is the number of times we must iteratively apply (33) until the absolute value of the result is less than one.

We call the sequence of values $\overrightarrow{C}_{r\circ n}(P)$, where $1 \leq n \leq \overrightarrow{C}_{r*}(P)$, the "conformal orbit" of P, which outlines a tetrational plot [72]. For example, the orbit with radix $r = 3$ of the quantum minus googol is $\{-10^{100}, -210.221, -5.49687, -2.09533, -1.05609, -0.096773\}$. No value of the orbit can be "identically 1" because ± 1 represents $\pm\infty$, while the global base b is our universe's maximum. Indeed, $\overrightarrow{C}_{e*}(b)$ gives us the universe's maximum natural depth.

We recover the original point by applying (34) iterated the same number of times, i.e.,

$$\overleftarrow{C}_{r\circ\overrightarrow{C}_{r*}(P)}\left(\overrightarrow{C}_{r\circ\overrightarrow{C}_{r*}(P)}(P)\right)$$

where $\overleftarrow{C}_{r\circ 1}(H) \equiv \overleftarrow{C}_{r}(H)$ and $\overleftarrow{C}_{r\circ(n+1)}(H) \equiv \overleftarrow{C}_{r}(H) \circ \overleftarrow{C}_{r\circ n}(H)$.

Every 1-ball of radius $\overleftarrow{C}_{r\circ n}(0)$ might correspond to a granularity level [73], a local setting belonging to the nested information of a (global) complex system such as the universe. Considering that $\overrightarrow{C}_{r*}(P)$ grows with P exceptionally slowly, the natural granularity levels are likely few; the ternary granularity depth for the currently estimated universe size in Planck units would be $\overrightarrow{C}_{3*}(10^{61}) = 5$, and the binary granularity depth for the currently estimated number of atoms in the known universe would be $\overrightarrow{C}_{2*}(10^{88}) = 8$.

Because every iteration conserves the local likelihood information, a granularity realm has an identical copy of the Bayesian data that matches its range of distances in the external world. Nevertheless, it is autonomous in creating new information elements, such as those resulting from clustering points or lumping together similarity classes, defining emerging organizational behavior. We can even take this combination of iterated coding with coarse-grained modeling [74] as a principle of multiscale modeling [75].

From a computational point of view, conformal coding might use a representation similar to the level-index number system [76]. The quantum $P \in \text{Å}^{+}$ encoded as the (true-normalized form of the) significand $0 < s < 1$ after $n = \overrightarrow{C}_{r*}(P) \geq 1$ iterations would be represented as $s_{r\circ n}$, so that

$$P \approx \overleftarrow{C}_{r\circ n}(s) = \frac{1}{2}\left(1 + r^{\frac{1}{2}\left(1+r^{\frac{1}{2}\left(1+r^{\cdots\frac{1}{2}\left(1+r^{s}\right)}\right)}\right)}\right)$$

where the order (height) of the power tower is n.

For example, a conformal representation of the number googol is $0.00212111221\ldots_{3\circ 5}$ (see Figure 9) owing to $0.00212111221\ldots_{3} \equiv 0.096773\ldots$ and

$$10^{100} = \overleftarrow{C}_{3\circ 5}(0.096773\ldots) = \frac{1}{2}\left(1 + 3^{\frac{1}{2}\left(1+3^{\frac{1}{2}\left(1+3^{\frac{1}{2}\left(1+3^{\frac{1}{2}\left(1+3^{0.096773\ldots}\right)}\right)}\right)}\right)}\right)$$

We similarly obtain that $10^{88} = \overleftarrow{C}_{2\circ 8}(0.805897\ldots) \equiv 0.1100111001\ldots_{2\circ 8}$.

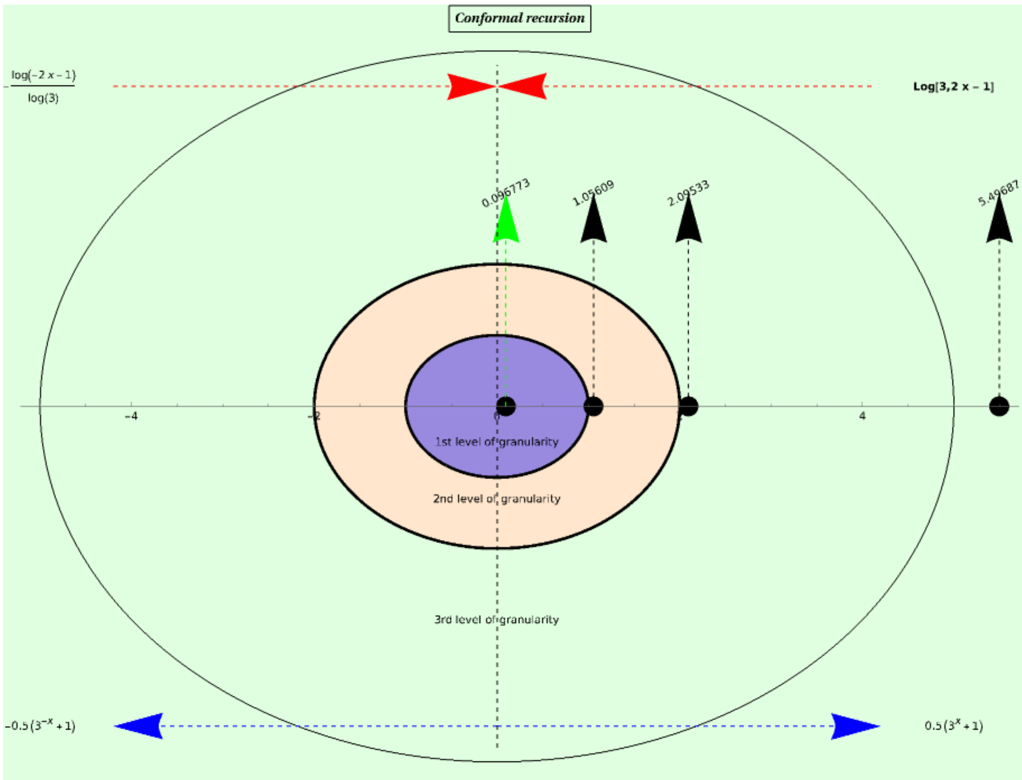

Figure 9. The three most profound levels of granularity $\left\{\pm\overleftarrow{C}_r(0)=\pm1,\pm\overleftarrow{C}_{r\circ2}(0)=\overleftarrow{C}_r(\pm1)=\pm2,\pm\overleftarrow{C}_{r\circ3}(0)=\overleftarrow{C}_r(\pm2)=\pm5\right\}$ (out of 5) the coding source in green generates to encode the number googol with radix $r=3$, the minor four points of the conformal orbit, and the conformal encoding (in red, Equation (33)) and decoding (in blue, Equation (34)) functions.

5. Primordial Distributions

We cannot irrefutably prove that the probability ISL (1) is a foremost PMF beating at the core of the cosmos, but NBL emanating from it is at least evidence supporting that possibility. Because NBL is pervasive and reflects the properties of PN, efficiency must be a rudimentary feature of nature. Following this trend of thought, we describe how to grow this one-parameter PMF from elemental geometric distributions, guaranteeing universal constructability. We will also tune the PMF's parameter to achieve the expected divisibility of $\mathbb{N}$'s probability mass and cast probability values as unit fractions. This provision is equivalent to making the event "picking the number one" a Bernoulli process.

The resulting PMF for the natural numbers allows calculating the probability and entropy of dichotomies like odd–even and prime–composite and trichotomies such as negative–zero–positive and elliptic–Euclidean–hyperbolic. The canonical PMF for the integer numbers could be the germ of a fundamentally unitary, parity-invariant, uncertain, discrete, and maximally entropic universal field.

5.1. Constructability

The Euler product formula for the zeta function (equations 1.6 and 1.13 in [77]) allows us to define PMF (1)'s primordial random variable $X \in \mathbb{N}^+$ as the infinite product

$$X=\prod_{p\in\mathbb{P}} p^{X(1/p^2)}$$

of independent identically distributed random variables with a geometric sequence of probabilities $\Pr\left(\underset{p\in\mathbb{P}}{X}(1/p^2)=k\right)=(1/p^2)^k(1-1/p^2)$ [40], corresponding to the PMF of k failures before the first success, with each binomial trial (see below) having a failure probability of the parameter $1/p^2$. This construction means that the canonical PMF (up to ϵ) consists of geometric distributions that, in turn, are "memoryless" [78] and "infinitely divisible" [79]. Of course, this infinitude is theoretical because there must be a maximum prime.

Once we have constructed the one-parameter canonical PMF from elemental generators, which constitutes a rudimentary notion of emergence, we can improve its computability by digging into rationality and making the probability mass of the natural numbers divisible on average to ensure its compartmentalization.

5.2. Ensuring Divisibility

At the end of Section 2.2, we determined that $\epsilon \in (0, 6/\pi^2)$ by merely demanding $\Pr(N \in \mathbb{N}) > 0$, i.e., $(1 - \epsilon\zeta(2)) = (1 - \epsilon\pi^2/6) > 0$. Given that the fraction of square grid $\mathbb{Z}^2$'s points visible from the origin is precisely $6/\pi^2$ [80], our requirement is equivalent to setting ϵ as a rational equal to the size of a subset of them.

We now introduce a new vital constraint on the number of divisors of a natural number intended to narrow the range of possible values of ϵ. The rationale is that divisibility, as a dual concept of primality, is paramount to understanding our universe.

Remember that a nonzero natural's number of divisors $đ$ includes one and the number itself; for instance, $đ(1^2) = 1$, $đ(2^2) = 3$ (1, 2, and 4), $đ(3^2) = 3$ (1, 3, and 9), and $đ(4^2) = 5$ (1, 2, 4, 8, and 16). Notwithstanding that the distribution $\Pr\left(đ(N^2)\right)$ is unknown, we can calculate the mean of divisors of counting numbers squared by employing the general law of the expected values (or "unconscious statistician") [81]. We require such an expected value to be at least two to guarantee that the divisibility of the entire probability room defined by (1) takes place non-trivially and naturally, for if N^2 splits into $đ(N^2) \geq 2$ parts, so can $\Pr(N)$, i.e.,

$$2 \leq \widehat{E}\left(đ\left(N^2\right)\right) = \epsilon \sum_{N=1}^{\infty} \frac{đ(N^2)}{N^2} = \epsilon \frac{\zeta^3(2)}{\zeta(4)} = \epsilon \frac{(\pi^2/6)^3}{\pi^4/90} = \epsilon \frac{5}{12}\pi^2$$

where the summation agrees with Equation (3.41) of [82]. Considering the high probability of picking the unit, this constraint is more rigid than it might seem at first sight.

Let us recap. We are imposing only a pair of constraints to provide ϵ with an accurate value; first, the probability of a natural to be nonzero, and second, the expected probability mass of the set of natural numbers to be splittable. Thus,

$$(0 < \epsilon < 6/\pi^2) \wedge \left(\epsilon \frac{5}{12}\pi^2 \geq 2\right)$$

constricts the possible values of ϵ to the narrow range

$$\epsilon \in \frac{1}{\pi^2}\left[\frac{24}{5}, 6\right) \approx (0.4863417, 0.6079271)$$

We aim to define NBL in a strict rational setting to increase operability through multiplication and division (see Sections 2.3 and 3.1). Although rationals such as $49/100$ or $6/10$ satisfy this constraint, the most probable numerator and denominator, in agreement with (1), are precisely 1 and 2, so

$$\epsilon = \frac{1}{2} \tag{39}$$

In passing, this value assures that the probability mass of a nonzero natural number is splittable. Moreover, it is a unit fraction.

5.3. Randomness

What does (39) mean from the information theory perspective? It means converting the event "picking the unit" into a Bernoulli (binomial) experiment equivalent to flipping a coin. A sequence of these independent identically distributed picks is a Bernoulli process, unique and universal in that it is the single most random non-mixing process possible [83].

More generally, the corresponding variable that considers obtaining exactly k ones in an experiment with N trials, each with a probability of success ϵ, follows the binomial distribution [84]

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

$$\Pr(k, N, \epsilon) = \binom{N}{k}\epsilon^k(1-\epsilon)^{(N-k)} \tag{40}$$

isomorphic to "exactly k tosses out of $N \geq k$ tosses resulting in a head". The logarithmic scale awesomely comes up again when we calculate the entropy of this distribution, which tends to $\frac{1}{2}\ln(2\pi e N\epsilon(1-\epsilon))$ as N approaches infinity.

Well, the binary entropy function of "picking the number one" (and the complementary event "picking a number with splittable probability") attains its maximum when (39) holds, like a fair coin, where the odds for and against are 1 (heads or tails) by definition. For the same reason, $\frac{1}{2}$ maximizes the entropy of the binomial distribution (40), too.

In summary, (39) ensures five conditions, to wit, positive rational probabilities congruent with the PMF itself, average divisibility of $\mathbb{N}$'s probability mass, maximum number of naturals with splittable occurrence probability, counting numbers with unit fraction probability masses, and maximum universal randomness.

5.4. Canonical PMF for the Natural Numbers

Assuming (39), we can establish that

$$\Pr(X = N) = \begin{cases} N \in \mathbb{N} - \{0\} : & \frac{1}{2N^2} \\ else : & 1 - \frac{1}{2}\zeta(2) \end{cases} \tag{41}$$

is the canonical PMF for a random variable X that takes natural values. Its mode is one, and its mean and variance are undefined. As required, the expected divisibility value of the global probability mass surpasses two, namely,

$$\widehat{E}\left(d\!\!\!/\left(N^2\right)\right) = \frac{1}{2}\frac{\zeta^3(2)}{\zeta(4)} \approx 2.05617$$

Indeterminacy has chances at every tick of the clock because of $1 - \frac{1}{2}\zeta(2) \approx 17.75\%$. Therefore, the probability of a counting number coming out is

$$\frac{1}{2}\sum_{N=1}^{\infty}\frac{1}{N^2} = \frac{\zeta(2)}{2} = \frac{\pi^2}{12} \approx 82.25\%$$

Let us see (42) in action. The probability of a string of numbers is the product of the individual probabilities; for example, the probability of picking s and t in a row is $\Pr(<s,t>) = \Pr(<t,s>) = \Pr(s)\Pr(t)$. The probability of a choice between a set of numbers is the sum of the individual probabilities; in particular, the probability of picking a number in the interval $[s..t)$ is the sum of $\Pr(s) + \Pr(s+1) + \cdots + \Pr(t-1)$.

The probabilities of a natural number being odd and even are $(3/4)(\zeta(2)/2) = (\pi/4)^2 \approx 5/8$ and $(1/4)(\zeta(2)/2) = \pi^2/48 \approx 5/24$, respectively (see Equation (1.12) in [85]). So, obtaining an odd natural number is $\frac{3/4}{1/4} = 3$ times as probable as picking an even, in sharp contrast with the "intuitive" $\frac{1}{2}$ size in [26].

The probability of the event "picking a natural number greater than 1", i.e., a number with splittable probability, is the unit minus the probabilities of indeterminate and one, namely,

$$\Pr(X \geq 2) = 1 - \left(1 - \frac{1}{2}\zeta(2) + \frac{1}{2}\right) = \frac{1}{2}(\zeta(2) - 1) \approx 0.32247$$

The probability of the event "picking a prime" is half the sum of the reciprocals of prime numbers $\mathbb{P}$ squared, i.e.,

$$\Pr(X = N \in \mathbb{P}) = \sum_{N \in \mathbb{P}} \frac{1}{2N^2} = \frac{1}{2}P_\zeta(2) \approx 0.22612$$

where $P_\zeta(\cdot)$ is the prime zeta function. Thus, picking a composite number has a probability of

$$\Pr(X \geq 2 \wedge X \notin \mathbb{P}) = \frac{1}{2}(\zeta(2) - 1 - P_\zeta(2)) \approx 0.09634$$

and the probability that $N \in \mathbb{P}$, conditioned to be greater than 1, exceeds 70% due to

$$\Pr(X = N \in \mathbb{P} | X \geq 2) = \frac{\Pr(X = N \in \mathbb{P})}{\Pr(X \geq 2)} = \frac{P_\zeta(2)}{\zeta(2) - 1} \approx \frac{0.22612}{0.32247} \approx 0.70123$$

Therefore, observing primes in nature is expected; regarding number theory, the odds of prime versus composite are $70/30$.

The canonical PMF (42) straightforwardly explains the supremacy of the hyperbolic configurations concerning the two-dimensional tilings algebraically associated with the finite reflection groups [86].

The probability of a natural number greater than one occurring thrice is

$$(\Pr(X \geq 2))^3 = \left(\frac{1}{2}(\zeta(2) - 1)\right)^3 \approx 0.033532$$

In other words, this value is the probability of producing three naturals to form a triangle.

The probability of picking three naturals forming a Euclidean triangle (i.e., $1/l + 1/m + 1/n = 1$) is the probability of picking $\{2, 3, 6\}$, $\{2, 4, 4\}$, or $\{3, 3, 3\}$, namely,

$$\left(\frac{1}{2}\right)^3 \left(\frac{1}{2^2 3^2 6^2} + \frac{1}{2^2 4^2 4^2} + \frac{1}{3^2 3^2 3^2}\right) \approx 0.000390$$

The probability of picking three naturals forming a spherical triangle (i.e., $1/l + 1/m + 1/n > 1$) is

$$\left(\frac{1}{2}\right)^3 \left(\frac{\zeta(2) - 1}{2^2 2^2} + \frac{1}{2^2 3^2 3^2} + \frac{1}{2^2 3^2 4^2} + \frac{1}{2^2 3^2 5^2}\right) \approx 0.005780$$

where the first term is the probability of a triplet of two 2s and any natural greater than one.

The probability of picking three naturals forming a hyperbolic triangle ($1/l + 1/m + 1/n < 1$) equals the occurrence probability of a triangle that is neither Euclidean nor spherical:

$$0.033532 - 0.000390 - 0.005780 \approx 0.0273615$$

So, the odds of hyperbolic cases against non-hyperbolic cases point to the former's predominance, specifically

$$\frac{0.0273615}{0.000390 + 0.005780} \approx 4.434429$$

In summary, two-dimensional tilings are usually hyperbolic. Therefore, inversive and conformal geometry might be at the world's heart, built from the simplest simplices (generalizations of a triangle to any dimension).

5.5. Canonical PMF for the Integer Numbers

$$\Pr(X = N) = \begin{cases} N \in \mathbb{N} - \{0\} : & \frac{1}{2N^2} \\ else : & 1 - \frac{1}{2}\zeta(2) \end{cases} \quad (42)$$

is the canonical PMF for a random variable X that takes natural values.

Finally, we can extend (42) to establish that

$$\Pr(\Xi = Z) = \begin{cases} Z \in \mathbb{Z} - \{0\} : & \frac{1}{(2Z)^2} \\ else : & 1 - \frac{1}{2}\zeta(2) \end{cases} \quad (43)$$

is the canonical PMF for a random variable Ξ that takes integer values. It has mode ± 1 (see Figure 10), with entropy

$$-\left(1 - \frac{1}{2}\zeta(2)\right)\log_2\left(1 - \frac{1}{2}\zeta(2)\right) - 2\sum_{Z=1}^{\infty}\frac{\log_2\left((2Z)^{-2}\right)}{(2Z)^2} = 3.44027 bit$$

For nonzero integers x and y, it is the solution to the functional equation

$$\Pr(x)\Pr(y) = (x\, y\, \Pr(x\, y))^2$$

satisfying the condition $\Pr(\pm 1) = \frac{1}{2}^2$.

This PMF, which accounts for the weight of the ISL in physics [87], cannot be more buildable; take every second integer from the coding source point of view, i.e., $\{\cdots, -8, -6, -4, -2, , 2, 4, 6, 8, \cdots\}$, then their reciprocals squared, obtaining $\{\cdots, 1/64, 1/36, 1/16, 1/4, , 1/4, 1/16, 1/36, 1/64, \cdots\}$, which coincides with (43). It satisfies positive probability masses summing to one, central reflection symmetry, fair (i.e., undefined) mean and variance, holistic rationality, and randomness. Unitarity, parity invariance, indeterminism, discreteness, and the principle of maximum entropy are the embodiments of these properties, which are all fundamental. In addition, the square root of the probability mass (physically, the probability amplitude) of a nonzero integer is precisely half its reciprocal.

Again, the resulting NBL is valid irrespective of the proportionality constant ϵ, following Section 2.5, namely,

$$\Pr(r, d) = \log_r\left(1 + \frac{1}{|d|}\right)$$

where $1 \leq |d| < r$, $r \in \mathbb{N} \geq 2$, and $d \in \mathbb{Z}$.

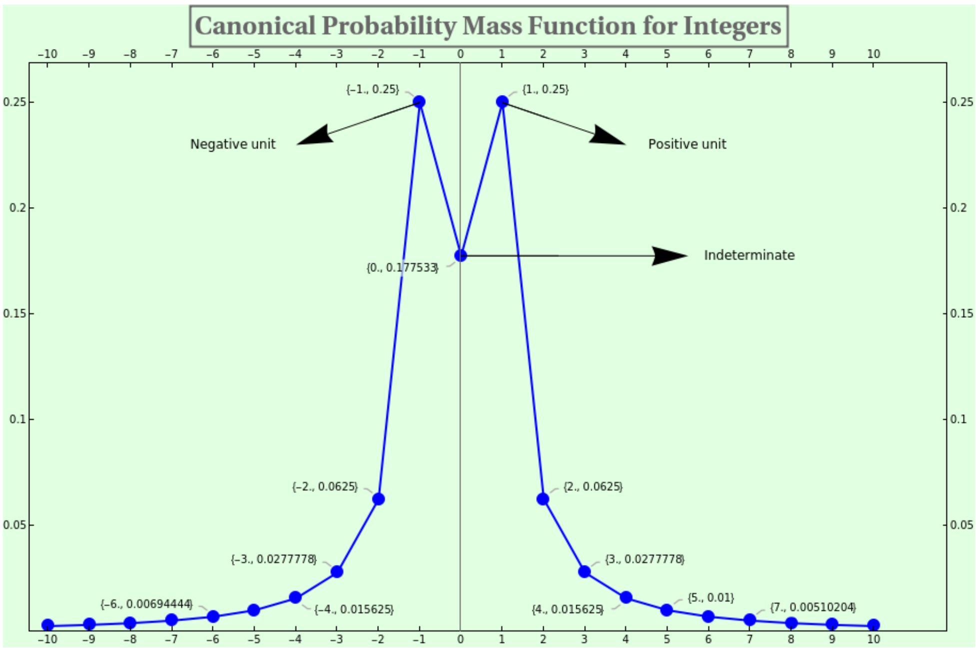

Figure 10. The canonical PMF for the integer numbers (43) could germinate a fundamentally unitary, parity-invariant, indeterministic, discrete, and maximally random universal field. The plot resembles the (inverted) cross-section of a sombrero potential (e.g., the quartic function $-(Z/2)^2 + 2(Z/2)^4$) of a scalar field with an unstable (indeterminate) center and a nonzero "vacuum expectation value"; the multiplicative units ∓ 1 provide this field with the ground (vacuum) state, enabling spontaneous symmetry breaking. Thus, the sombrero potential would be the physical manifestation of a fundamental improbability mass function.

6. Conclusions

Our research shows how discreteness and the continuum interact under the shelter of (1). A complex system and its environment embody the continuous local and discrete global. The NBL probability's derivative of the local takes us to the global, and vice versa; the global's integral situates us in a local setting of likelihood-based probability. A harmonic scale of rational numbers supports the global realm, while a logarithmic scale of "real values" supports the local realm. The harmonic scale's base confines the rational setting, and a logarithmic scale's radix is an exponentiation constant that normalizes a complex system's conformal space of "continuous" information coded in PN.

We consider a complex system a coding source that observes the outside, operates internally with the gathered information, and takes action on the environment. More precisely, a coding source uses the synergy between Benford's and Bayes's laws to reflect (encode) the external world, process (recode or arithmetically transform) the information, and return (decode) the results to its immediate surroundings. These laws connect mathematics with physics.

6.1. Canonical PMF

NBL bets on smallness; little objects are more numerous than extensive ones. Why? The fact that many probability distributions partially adhere to NBL does not reveal its root. Nor can we glean its origin from the fact that merging methods via sampling or multiplication of real-world data series produces adherence to NBL. "Mathematics alone cannot justify a first-digit law", wrote Raimi. Given that the effects of NBL are well known in physics, we need to be aware of its fundamental character and ultimate cause.

NBL is not so mysterious if we concede that it originates from an ISL of probability. Whenever the canonical PMF governs a system behavior, we can infer its properties are data spaces that record information on a positional scale. This constructible primordial PMF states an absolute hyperbolic relation between the square of a nonzero number $N \in \mathbb{N}$ and its probability mass $\Pr(N)$; $N^2 \Pr(N)$ is constant.

If we assume the minimal information principle, PMF (43) is the only way to satisfy that the positive probabilities sum to 1, guarantee that our random variable's average value is not finite if we repeat the experiment often enough, ensure mass divisibility, and cope with the extension to all the integers. These requirements' logic, sturdiness, and feasibility suggest a full-fledged tenet at the heart of mathematics, physics, and higher integrative levels. In addition, this PMF implements the Axiom of Induction; "at least one inductive set does exist determined uniquely by its members that has a substrative probability distribution".

Although the canonical PMF consists of Bernoulli generators, we introduce it as a brute fact deprived of tangible information. This probability ISL could be the embryo for crucial experimental laws of physics, such as Newton's universal gravitation and Coulomb's electrostatic force. Understood as an improbability field, it could even give place to the energy density of free space characterized by a sombrero potential with an unstable center and a nonzero vacuum expectation value.

The canonical PMF indicates a manner of arranging availability versus transcendence. Occasional numbers are more startling and influential than abundant ones. While frequent numbers provide resilience, infrequent numbers have the capacity for transformation. Therefore, the universal equilibrium is not enforced via uniformity but achieved by hyperbolically balancing accessibility or stability (position) against magnitude or reactivity (momentum); does this not sound like the uncertainty principle?

The canonical PMF naturally copes with ambiguity and vagueness by introducing the "indeterminate" value (interpreted as inaction or not-a-number and symbolized by zero) and dodging a finite expected value. However, pure conformality is only possible within a finite global scope; b ultimately enables implementing a PN-based coding space where the accessibility potential

$$\Pr(b,q) + q\frac{\partial \Pr(b,q)}{\partial q}$$

vanishes for all quanta q and the logarithm can germinate.

6.2. The Logarithm Measures Local Information

Integrating under the global NBL's hyperbola and normalizing concerning a local radix $r \ll b$ immediately drives us to the fiducial, $\mathbb{R}$-based, logarithmic form of NBL.

The logarithm and its inverse (the exponential) are fundamental functions because they appear everywhere in mathematics and physics. The logarithm bridges information and physics, especially thermodynamics, via the Gibbs entropy formula in statistical mechanics. In information theory, the logarithm mainly estimates the representational extent of a given numeral written in PN. Further, this manuscript proves that the logarithm resolves the metric of a conformal space by recasting correlations into distances or rapidities. This conversion is critical in iterated coding, especially in coarse-grained and multiscale modeling. The radix's logarithm is precisely the (absolute value of the) coding space's curvature. In particular, radix $r = \mathrm{e}$ defines the natural logarithmic scale, i.e., the standard one-dimensional hyperbolic space.

The logarithm is central to comprehending how profoundly NBL connects with recurrence and incrementality. A coding source implements Bayes' rule by multiplying the prior odds between two quanta by a likelihood factor that is precisely the logarithm of their ratio's reciprocal. This structure represents likelihood information, e.g., the encoded odds of an elemental jump, and admits a representation as a gap ratio $(A-B)/(C-D)$, which is recurrent under iterative processes of encoding, recoding, arithmetic operations, and decoding.

When a digit is one of the $r-1$ available items in a pool, Bayesian coding solves a version of "the secretary problem" that considers the strategy to select "a good" item rather than "the best" one. It belongs to a class called "last-success problems" with universal scope. Its objective is to determine the last item x on the fly that maximizes the probability of success in accomplishing the stopping rule, i.e., rejecting the first x items and stopping afterward at the next that is better than the preceding ones. We approach the solution by aggregating the past odds $\tilde{o}(x:1|r) \propto \frac{r}{x}\ln x$ and the future odds $\tilde{o}(r:x|r) \propto x\ln\frac{r}{x}$. We feature a couple of characteristic properties of the local likelihood distribution function (bipartite entropy) that solves these good-choice (best-choice included) and optimal-stopping problems; first, it has a phase for incremental information gathering to deliver a preliminary output (past maximum) only refined (future maximum) if there is time left, and second, the solution x is generally asymmetric, in that it makes the past partition (bin $[1,x)$) more minor than the future partition (bin $[x,r)$), i.e., information flows towards the origin.

6.3. Thrifty World

NBL denotes productivity. Radix economy $\mathring{E}(N,r) \approx r\log_r N$ measures the "price" of a numeral N using radix r as a parameter. Cost-saving number systems will employ an efficient coding radix; the optimal radix economy corresponds to Euler's number e, another sign of the preeminence of small numbers. The wider the gap between the economy of consecutive numbers relative to the radix, the higher the expected frequency. Thrifty numbers making a difference are winning, meaning that the probability of a number coded with radix r showing up is the rate of change, or derivative, of its economy concerning the radix, specifically

$$P(N,r) = \frac{\mathring{E}(N+1,r) - \mathring{E}(N,r)}{r} \approx \log_r(N+1) - \log_r N = \log_r\left(1+\frac{1}{N}\right)$$

This expression indicates the occurrence probability of the numeral N, not necessarily a digit, with radix r. For example, $\log_{10}(1+1/22)$ is the probability of running into a decimal number starting with 22, such as 2.29 or 2237. The logarithmic scale shirrs the surrounding space of a coding source; the closer, the lesser the numeral density. A large numeral is less likely due to its representational magnitude, so the space around it is tighter than that occupied by a numeral with more probability mass. NBL reflects how PN encodes numerals in agreement with this economic criterion.

Therefore, the radix economy establishes a scalar field where the gap between the "potential energies" of two objects only depends on their position as perceived from the source. Thus, the canonical PMF and NBL subsidiaries are fundamentally efficient, balancing probability mass against notation size. Minor numbers are accessible at a lower cost, while likelihood information is increasingly less available as we climb to infinity.

Our theory also associates efficiency with entropy. We can interpret NBL probabilities as degrees of stability or coherence. The lowest digits maintain distinctness from the surroundings thanks to their solid entropic support. The more significant digits are vulnerable and give rise to more transitions, i.e., higher reactivity or less resistance to integration with the environment.

Parsimonious management of computational resources is crucial, as optimal stopping problems reveal. In "the secretary problem", selecting "the best" applicant is pragmatically less sensible than simply "a good" one, which requires maximizing the bipartite entropy. We split the total likelihood provided by the local Benford–Bayes rule. The "past partition" emphasizes the information gathered, while the "future partition" deals with the information we can obtain from forthcoming aspirants. As the number of examined applicants grows, future information becomes past information, while the usage probability transfers oppositely. The best applicant implies focusing on the future partition exclusively, but a balanced decision also implies contemplating the past. Overall, the bipartite entropy and the solution are biased toward the origin and the past, respectively. This primary memory allocation in real-time is a default resource management strategy.

Information economy streamlines calculations. That the universe optimizes computability follows from NBL embracing several invariances. Base invariance ensures even interaction with the environment because changes in the radix value imply only incremental updates (recoding), keeping the internal metric up to the curvature. Scale invariance provides the means to recursively operate on nested levels of domain granularity, like a fractal. Ultimately, PN is effective because it makes the most expected data readily accessible for iterative coding functions.

6.4. Relationalism

Literature about NBL overlooks its rational aspect. Real numbers are unattainable mathematical objects, mere abstractions. In contrast, relative odds, i.e., proportions between two numbers, quanta, or digits, are tractable. Rational numbers fit in an inaccurate world where relations are as important as individual entities and comparative quantities predominate over absolute values. A universe built upon $\mathbb{Q}$ facilitates divisibility, discreteness, and operability.

Ours is primarily a harmonic world, where rational numbers guarantee calculability. PMF (42) tells us that the probability of nonzero natural numbers is a unit fraction, ensuring the average divisibility of the entire probability space. The global NBL, stemming from (1), means that quantum frequencies are unit fractions, too. The continuous (local, real) NBL emanates precisely from the discrete (global, rational) NBL by compartmentalizing a one-dimensional hyperbolic space of colossal extent.

The concept of information is fundamentally rational. The harmonic likelihood is global information defined as $[\mathcal{L}(q)]_s^t \equiv {}^{[\psi(q)]_s^t}/_{[\psi(q)]_1^2}$harmt, whereas logarithmic likelihood is local information defined as $[\ell(d)]_i^j \equiv {}^{[\ln d]_i^j}/_{[\ln d]_1^e}$nat. A "harmt" is the global (harmonic) unit of information, peering the local (logarithmic) unit of information, the "nat". Likewise, NBL describes normalized information regarding the global base b, $\Pr(b,[s,t)) \equiv {}^{[\mathcal{L}(q)]_s^t}/_{[\mathcal{L}(q)]_1^b}$, or the local radix, $\Pr(r,[i,j)) \equiv {}^{[\ell(d)]_i^j}/_{[\ell(d)]_1^r}$. The likelihood is space on a harmonic (global) or logarithmic (local) scale; for example, if we assume that the "bit" (2^1 possible states) is the minimal (unit) length, one "byte" (2^8 possible states) is eight units long. If our world was positional, likelihood and entropy would have metric units of length for all practical purposes, meaning that information would be a physical and manageable resource.

The rational character of our universe pops up in all its splendor when we address probability ratios. A source encodes, recodes, and decodes odds using Bayes' law in odds form. The odds value between a pair of numbers, quanta, or numerals is the quotient of their picking probabilities, quantifying the strength of their association. Assuming $a < b$, ${}^{\Pr(a)}/_{\Pr(b)} > 1$ estimates how uncorrelated a and b are; if ${}^{\Pr(a)}/_{\Pr(b)} \approx 1$, both events are mutually dependent. The global Bayes rule says that the odds of quantum s against t in base b are the odds of the number s versus t times the probability of the bucket $[s,t)$ in base b. We derive from this the stability of a quantum (16). The local Bayes rule says that the odds of digit i against j with radix r are the global odds of the quantum i versus j times the probability of the bin $[i,j)$ with radix r. We derive from this the stability of a digit (17). Both represent the entropic contribution of the items in a range to a positional scale, confirming that information is relational.

An exceptional case of Bayesian data is the cross-ratio, a conformality invariant. Despite the conformal coding functions using the logarithm and the exponential function, power (infinite) series by definition, the coding source adds or multiplies incrementally a finite series of cross-ratio powers to throw a rational result at any time (Equations (36) and (37)), bettering the approximation with the number of iterates. Rationality intricately intertwines with decidability in polynomial time and interruptible algorithms.

Numeric values do not contain information per se, while a common property makes two entities commensurable, with the global base and the local radix as main referents. We can take global Bayesian data as rational quanta, computable numbers, and local Bayesian data as observable correlations of numerals. More fundamentally, holistic rationality implies relationalism (i.e., reciprocity) and operability (i.e., arithmetical tractability), basic properties of a physical transformation. Rationality is budding relativity. Noticeably, Section 4.3 challenges a premise of special relativity; the global base, interpreted as the speed of light, proves the universe's "rational" rather than "real" essence. Indeed, the quotidian continuum we perceive from our local outlook approximates the discrete reality; the continuum emerges from the rational.

6.5. Conjectures

The canonical PMF, NBL subsidiaries, and Bayes' rules explain why proximity or slightness provides more stability than distance or heftiness. Like gravity, nature builds physics upon proximity because occurrence probability attracts information toward a central source. However, the entropic leaks from the most outlying digits offset this mass accumulation in the source's immediacy. Probability density dilutes as we move away from the origin, which resembles the second thermodynamical law. Therefore, data encoded in PN would induce alternating uphill and downhill flows, reflecting a brute fluctuation between the dual elementary concepts of concentration and dispersion. Our Benford–Bayes laws allow for the inference of this fundamental dynamic.

If the integer line $\mathbb{Z}$ were the position space of a generic object, the canonical PMF for the integers (43) would match with a default wave function with one degree of freedom, expressible as a linear combination of the position eigenstates $|Z\rangle$, namely,

$$\Psi = \cdots + {}^{1}\!/\!{}_{4}\,|-2\rangle + {}^{1}\!/\!{}_{2}\,|-1\rangle + \sqrt{1 - \frac{1}{2}\zeta(2)}\,|0\rangle + {}^{1}\!/\!{}_{2}\,|+1\rangle + {}^{1}\!/\!{}_{4}\,|+2\rangle + {}^{1}\!/\!{}_{6}\,|+3\rangle + \cdots$$

A nonzero integer $Z \in \mathbb{Z}$ represents an actual eigenvalue corresponding to the eigenstate with rational amplitude $(2Z)^{-1}$, and the origin is the unstable (beable) central state with quantum amplitude $\sqrt{1 - \frac{1}{2}\zeta(2)}$. By the Born rule, the quantum-mechanical probability of being at place Z is the square modulus of its rational amplitude, precisely its canonical probability mass. From a complementary point of view, the canonical PMF tells us the probability $\Pr(Z) = (4E|Z\rangle)^{-1}$ of having energy $E|Z\rangle = Z^2$, corresponding to a wave function's rational amplitude $\left(2\sqrt{E|Z\rangle}\right)^{-1}$ in the momentum space.

The Benford–Bayes laws regulate the implementation of a conformal space through tractable hyperbolic functions. A Bayesian datum is a gap ratio, a particular case of gap ratios is a cross-ratio, and the logarithm of a cross-ratio yields a conformal metric. The canonical coding functions define how a complex system, say a coding source, creates an image of the world, which can render crucial consequences in physics, principally implementing scalability and boosting efficiency as leitmotifs and chief drivers of cosmological development.

Infinity has no place in the algebraic field of rational numbers, so it lacks physicality. We have even factually inferred that information divergence is impossible. In the first place, the entropy of the canonical PMF for the natural and integer numbers converges. Then, we feature a global base closely related to the maximum natural (or prime) number. Likewise, a ball with radius $1 - 1/b$ confines the local Bayesian data. The jump odds between consecutive quanta or numerals are also delimited. Physically, the entropic cost of crossing entirely the universe or its local copy agrees with the Bekenstein bound. Perceivable things in the cosmos are typically small but always rationally commensurable from some standpoint; otherwise, they would be incomparable, thus indiscernible. An infinite host universe would reduce all its finite guests to zero, an unobservable number. Moreover, a transfinite universe prone to productivity is counterintuitive. The universe is an economic system precisely owing to its limited scope and resources.

This investigation generally points to mathematics having a potential physical status. We have put laws midway between mathematics and physics on the table. The canonical PMF for integer numbers defines a pervasive numeric field of stability that is the germ of a constitutively unitary, parity-invariant, uncertain, discrete, and maximally random universe. The linkage between probability space and physical space, especially within a coding space, is so intricate that we hardly find discrepancies. Further, because the notion of logarithmic likelihood results from comparing two logarithmic sectors (5) and a local NBL probability mass is a ratio of logarithmic likelihoods (6), stating that information is physical means probability is physical.

We have told a hegira from information coding to physics, presuming the Galilean idea that nature is mathematical per se. William K. Clifford (1976, On the Bending of Space, https://doi.org/10.1007/978-94-010-1727-5) underlined that we might "be treating merely as physical variations effects which are really due to changes in the curvature of our space", although "Whether one associates 'geometric' ideas with a theory is [illegible] a private matter", stated Einstein in a letter written to Reichenbach (Google translation from Doc. AEA 20-117 of the Albert Einstein Archive, 8 April 1926). Our partway philosophical worldview supports the theory that physics emerges from algebra via geometry, supported by hyperbolicity, economy, and relationalism. The embodiment of these pillars makes Tegmark's hypothesis that the observable reality is a mathematical structure defined by computable functions plausible. We must add that such a structure consists of conformal spaces and transformations.

Funding: This research received no external funding

Data Availability Statement: The original contributions presented in this study are included in the article. The author can be contacted for further inquiries. The programming code producing the data presented in this study is available on request.

Acknowledgments: We herein express our tribute to some personalities and entities influencing this work. The UNED has allowed the development of a scientific instinct and a generalistic approach to problem solving. Working for Indra enabled a pragmatic approach to figure out any crisis and awareness of the productive character of our universe. We greatly thank José Mira, Joan Baez, Lee Smolin, Sean Carroll, Norman J. Wildberger, and Sabine Hossenfelder for their wisdom and explanatory capacity. We apologize in advance for omitting laudable references. We sincerely appreciate those offering comments and critical insight on this version.

Conflicts of Interest: The author asserts that only scientific rigor, significance, and clarity drive the high-level goals of this work. It contains no known minor or significant incongruencies, errors, or inaccuracies. The author did not receive support from any organization for the submitted work and declares that he has no competing financial or non-financial interests directly or indirectly related to the work submitted for publication. No personal relationships have influenced the content of

this work. This work is an original one-piece that must be kept unsplit into several parts. It has not been published elsewhere in any form or language, partially or in whole. The author claims to have committed no intentional ethical wrongdoing related to this paper, including self-plagiarism, plagiarism, far-fetched self-citations, conflict of interest, inaccurate authorship declarations, and unacceptable biases concerning the references. This work respects third-party rights such as copyright and moral rights.

References

1. Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. *Am. J. Math.* **1881**, *4*, 39–40.
2. Benford, F. The Law of Anomalous Numbers. *Proc. Am. Philos. Soc.* **1938**, *78*, 551–572.
3. Hürlimann, W. Benford's law from 1881 to 2006. *Res. Gate* **2006**. http://arxiv.org/abs/math/0607168.
4. Blair, D.E. *Inversion Theory and Conformal Mapping*; Graduate Studies in Mathematics; American Mathematical Society: Providence, RI, USA, 2000.
5. Burke, J.; Kincanon, E. Benford's Law and Physical Constants: The Distribution of Initial Digits. *Am. J. Phys.* **1991**, *59*, 952–952.
6. Berger, A.; Hill, T. Benford's Law Strikes Back - No Simple Explanation in Sight for Mathematical Gem. *Math. Intell.* **2011**, *33*, 85–91. https://doi.org/10.1007/s00283-010-9182-3.
7. Berger, A.; Hill, T.P. What is Benford's Law? *Not. AMS* **2017**, *64*, 132–134. https://doi.org/10.1090/noti1477.
8. Hossenfelder, S. *Lost in Math: How Beauty Leads Physics Astray*; Basic Books: New York, NY, USA, 2018.
9. Finch, S.R. *Mathematical Constants II*; Encyclopedia of Mathematics and Its Applications; Cambridge University Press: Cambridge, UK, 2018.
10. Witten, E. A Mini-introduction to Information Theory. *Riv. Del Nuovo C.* **2020**, *43*, 187–227. https://doi.org/10.1007/s40766-020-00004-5.
11. Bayes, T. LII. An Essay Towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philos. Trans. R. Soc. Lond.* **1763**, *53*, 370–418.
12. Oldershaw, R.L. Nature is Startling—Conformal Symmetry. *Medium* **2017**. https://doi.org/10.13140/RG.2.2.12399.74407.
13. Formann, A.K. The Newcomb-Benford Law in Its Relation to Some Common Distributions. *PLoS ONE* **2010**, *5*, e10541. https://doi.org/10.1371/journal.pone.0010541.
14. Ariew, R. *Ockham's Razor: A Historical and Philosophical Analysis of Ockham's Principle of Parsimony*; University of Illnois at Urbana-Champaign: Champaign, IL, USA, 1976.
15. Planck Collaboration; Aghanim, N.; Akrami, Y.; Ashdown, M.; Aumont, J.; Baccigalupi, C.; Ballardini, M.; Banday, A. J.; Barreiro, R. B.; Bartolo, N.; et al. Planck 2018 results—VI. Cosmological parameters. *Astron. Astrophys.* **2020**, *641*, A6. https://doi.org/10.1051/0004-6361/201833910.
16. Helmenstine, A. Composition of the Universe—Element Abundance. Science Notes. 2022. Available online: https://sciencenotes.org/composition-of-the-universe-element-abundance/ (accessed on 30 October 2024).
17. Mo, H.; van den Bosch, F.; White, S. *Galaxy Formation and Evolution*; Galaxy Formation and Evolution; Cambridge University Press: Cambridge, UK, 2010.
18. Fewster, R.M. A Simple Explanation of Benford's Law. *Am. Stat.* **2009**, *63*, 26–32.
19. Tegmark, M. The Mathematical Universe. *Found. Phys.* **2008**, *38*, 101–150. https://doi.org/10.1007/s10701-007-9186-9.
20. McCreary, P.R.; Murphy, T.J.; Carter, C. The Modular Group. *Math. J.* **2018**, *20*, 1–33. https://doi.org/10.3888/tmj.20-3.
21. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351.
22. Berger, A.; Hill, T.P.; Rogers, E. Benford Online Bibliography. 2009. Available online: http://www.benfordonline.net (accessed on 21 February 2023).
23. Miller, S.J. *Benford's Law*; Princeton University Press: Princeton, NJ, USA, 2015.
24. Berger, A.; Hill, T.P. *An Introduction to Benford's Law*; Princeton University Press: Princeton, NJ, USA, 2015.
25. de Finetti, B. *Probability, Induction, and Statistics: The Art of Guessing*; Wiley Series in Probability and Statistics; John Wiley: New York, NY, USA, 1972.
26. Hill, T.P. A Statistical Derivation of the Significant-Digit Law. *Stat. Sci.* **1995**, *10*, 354–363.
27. Havil, J. *Gamma—Exploring Euler's Constant*; Princeton University Press: Princeton, NJ, USA, 2003.
28. Hill, T.P. The Significant-Digit Phenomenon. *Am. Math. Mon.* **1995**, *102*, 322–327. https://doi.org/10.1080/00029890.1995.11990578.
29. Hill, T.P. The First Digit Phenomenon. *Am. Sci.* **1998**, *86*, 358–363.
30. Tao, T. Benford's Law, Zipf's Law, and the Pareto Distribution. Terrytao's Blog. 2009. Available online: https://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/ (accessed on 30 October 2024).
31. Simon, H.A. On a Class of Skew Distribution Functions. *Biometrika* **1955**, *42*, 425–440.
32. Fang, G.; Chen, Q. Several Common Probability Distributions Obey Benford's Law. *Phys. A Stat. Mech. Its Appl.* **2020**, *540*, 123129. https://doi.org/10.1016/j.physa.2019.123129.

33. Berger, A.; Hill, T.P. The Mathematics of Benford's Law: A Primer. *Stat. Methods Appl.* **2021**, *30*, 779–795. https://doi.org/10.1007/s10260-020-00532-8.
34. Gonsalves, R.A. Benford's Law—A Simple Explanation. *Medium* **2020**.
35. DK. A Treatise On The Newcomb-Benford Law. *Medium* **2021**.
36. Wikipedia Contributors, the Free Encyclopedia. Zipf's Law, 2022. Available online: https://simple.wikipedia.org/wiki/Zipf%27s_law (accessed on 1 April 2022).
37. Hendry, R.F. Structure, Scale and Emergence. *Stud. Hist. Philos. Sci. Part A* **2021**, *85*, 44–53. https://doi.org/10.1016/j.shpsa.2020.08.006.
38. Williams, P.M. Bayesian Conditionalisation and the Principle of Minimum Information. *Br. J. Philos. Sci.* **1980**, *31*, 131–144.
39. Montemurro, M.A. Beyond the Zipf-Mandelbrot Law in Quantitative Linguistics. *Phys. A Stat. Mech. Its Appl.* **2001**, *300*, 567–578. https://doi.org/10.1016/S0378-4371(01)00355-7.
40. Wikipedia Contributors, the Free Encyclopedia. Zeta Distribution, 2021. Available online: https://en.wikipedia.org/wiki/Zeta_distribution(accessed on 4 April 2022).
41. Finch, S.R. *Mathematical Constants*; Encyclopedia of Mathematics and Its Applications; Cambridge University Press: Cambridge, UK, 2003.
42. Held, L.; Bové, D.S. *Applied Statistical Inference: Likelihood and Bayes*; Springer: Berlin Heidelberg, Germany, 2013.
43. Bak, P.; Tang, C.; Wiesenfeld, K. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.* **1987**, *59*, 381–384. https://doi.org/10.1103/PhysRevLett.59.381.
44. Szendro, P.; Vincze, G.; Szasz, A. Pink-noise Behaviour of Biosystems. *Eur. Biophys. J.* **2001**, *30*, 227–231. https://doi.org/10.1007/s002490100143.
45. Mattey, G.J. *Lecture Notes on "Critique of Pure Reason": Appearance and Thing in Itself*; Philosophy 175 Home Page; UC Davis Philosophy Department: Davis, CA, USA, 1997.
46. Stone, J.V. *Information Theory: A Tutorial Introduction*; Tutorial Introduction Book, Tutorial Introductions; 2015.
47. Fechner, G.T.; Boring, E.G.; Adler, H.E.; Howes, D.H. *Elements of Psychophysics*; Henry Holt Edition in Psychology; Holt, Rinehart and Winston: New York, NY, USA, 1966; Volume 1.
48. Schatte, P. On Benford's Law to Variable Base. *Stat. Probab. Lett.* **1998**, *37*, 391–397. https://doi.org/https://doi.org/10.1016/S0167-7152(97)00142-9.
49. Howson, C., The Logic of Bayesian Probability. In *Foundations of Bayesianism*; Corfield, D.; Williamson, J., Eds.; Springer: Dordrecht, The Netherlands, 2001; Volume 24, pp. 137–159. https://doi.org/10.1007/978-94-017-1586-7_6.
50. Joyce, J. Bayes' Theorem. In *The Stanford Encyclopedia of Philosophy*, Spring ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2019.
51. Walker, S.; Cronin, L. Time Is an Object. Aeon. 2023. Available online: https://www.oasis.cnrs.fr/news/aeon-time-is-an-object (accessed on 30 October 2024).
52. Wikipedia Contributors, the Free Encyclopedia. Odds Algorithm, 2021. Available online: https://en.wikipedia.org/wiki/Odds_algorithm(accessed on 4 June 2021).
53. Girdhar, Y.; Dudek, G. Optimal Online Data Sampling or How to Hire the Best Secretaries. In Proceedings of the Canadian Conference on Computer and Robot Vision, Kelowna, BC, Canada, 25–27 May 2009; pp. 292–298. https://doi.org/10.1109/CRV.2009.30.
54. Wikipedia Contributors, the Free Encyclopedia. Golomb Coding, 2024. Available online: https://en.wikipedia.org/wiki/Golomb_coding(accessed on 12 December 2024).
55. Milne, J.J. *An Elementary Treatise on Cross-Ratio Geometry: With Historical Notes*; The University Press: Cambridge, UK, 1911.
56. Wikipedia Contributors, the Free Encyclopedia. Doppler Effect, 2024. Available online: https://en.wikipedia.org/wiki/Doppler_effect(accessed on 4 February 2024).
57. Tipler, P.A.; Llewellyn, R. *Modern Physics*; W. H. Freeman: New York, NY, USA, 2012.
58. Krakauer, D. Complexity, Agency, and Information. Sean Carroll's Mindscape. 2023. Available online: https://www.preposterousuniverse.com/podcast/2023/07/10/242-david-krakauer-on-complexity-agency-and-information/ (accessed on 30 October 2024).
59. Young, N.J. Linear fractional transformations in rings and modules. *Linear Algebra Its Appl.* **1984**, *56*, 251–290. https://doi.org/https://doi.org/10.1016/0024-3795(84)90131-9.
60. Stillwell, J. Modular Miracles. *Am. Math. Mon.* **2001**, *108*, 70–76. https://doi.org/10.1080/00029890.2001.11919727.
61. Poincaré, H. *Science and Hypothesis (1905)*; Read Books Limited: Redditch, UK, 2016.
62. Wikipedia Contributors, the Free Encyclopedia. Poincaré Disk Model, 2023. Available online: https://en.wikipedia.org/wiki/Poincar%C3%A9_disk_model(accessed on 20 January 2023).
63. Wikipedia Contributors, the Free Encyclopedia. Real Projective Line, 2023. Available online: https://en.wikipedia.org/wiki/Real_projective_line(accessed on 29 January 2024).
64. Trillas, E. Sobre Funciones de Negacion en la Teoria de Conjuntos Difusos. *Stochastica* **1979**, *3*, 47–60.

65. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*; Applied Mathematics Series; Dover Publications: Mineola, NY, USA, 1965.
66. Richeson, D.S. *Euler's Gem: The Polyhedron Formula and the Birth of Topology*; Princeton Science Library; Princeton University Press: Princeton, NJ, USA, 2019.
67. Karapetoff, V. Restricted Theory of Relativity in Terms of Hyperbolic Functions of Rapidities. *Am. Math. Mon.* **1936**, *43*, 70–82.
68. Russell, N. Framing Lorentz Symmetry. CERN Courier. 2004. Available online: https://cerncourier.com/a/framing-lorentz-symmetry/ (accessed on 30 October 2024).
69. Jaynes, E.T. Information Theory and Statistical Mechanics. In *Statistical Physics*; Ford, K.W., Ed.; Benjamin: New York, NY, USA, 1963.
70. Wikipedia Contributors, the Free Encyclopedia. Differential Entropy, 2024. Available online: https://en.wikipedia.org/wiki/Differential_entropy(accessed on 10 September 2024).
71. Bekenstein, J.D. Black Holes and the Second Law. *Lett. Nuovo C. (1971–1985)* **1972**, *4*, 737–740. https://doi.org/10.1007/BF02757029.
72. Wikipedia Contributors, the Free Encyclopedia. Tetration, 2024. Available online: https://en.wikipedia.org/wiki/Tetration(accessed on 31 January 2024).
73. Maier, J.F.; Eckert, C.M.; Clarkson, P.J. Model Granularity in Engineering Design, Concepts and Framework. *Des. Sci.* **2017**, *3*, e1. https://doi.org/10.1017/dsj.2016.16.
74. Seiferth, D.; Sollich, P.; Klumpp, S. Coarse Graining of Biochemical Systems Described by Discrete Stochastic Dynamics. *Phys. Rev. E* **2020**, *102*, 062149. https://doi.org/10.1103/physreve.102.062149.
75. Weinan E. *Principles of Multiscale Modeling*; Cambridge University Press: Cambridge, UK, 2011.
76. Clenshaw, C.W.; Olver, F.W.J. Beyond Floating Point. *J. Assoc. Comput. Mach.* **1984**, *31*, 319–328.
77. Peltzer, A.R. The Riemann Zeta Distribution. Ph.D. Thesis, University of California, Irvine, CA, USA, 2019.
78. Mitzenmacher, M.; Upfal, E. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*; Cambridge University Press: Cambridge, UK, 2017.
79. Ken-Iti, S. *Lévy Processes and Infinitely Divisible Distributions*; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 1999.
80. Castellanos, D. The Ubiquitous Pi. *Math. Mag.* **1988**, *61*, 67–98.
81. Billingsley, P. *Probability and Measure*; Wiley Series in Probability and Statistics; Wiley: Hoboken, NJ, USA, 1995.
82. Mathar, R.J. Survey of Dirichlet Series of Multiplicative Arithmetic Functions. *arXiv* **2011**, arXiv:1106.4038.
83. Ornstein, D. Bernoulli Shifts with the Same Entropy Are Isomorphic. *Adv. Math.* **1970**, *4*, 337–352. https://doi.org/10.1016/0001-8708(70)90029-0.
84. Sinharay, S. Discrete Probability Distributions. In *International Encyclopedia of Education*, 3rd ed.; Peterson, P., Baker, E., McGaw, B., Eds.; Elsevier: Oxford, UK, 2010; pp. 132–134. https://doi.org/10.1016/B978-0-08-044894-7.01721-8.
85. Alzer, H.; Choi, J. Four Parametric Linear Euler Sums. *J. Math. Anal. Appl.* **2020**, *484*, 123661. https://doi.org/10.1016/j.jmaa.2019.123661.
86. Grove, L.C.; Benson, C.T., Finite Groups in Two and Three Dimensions. In *Finite Reflection Groups*; Springer: New York, NY, USA, 1985; pp. 5–26. https://doi.org/10.1007/978-1-4757-1869-0_2.
87. Witten, E. Magic, Mystery, and Matrix. *N. Am. Math. Soc.* **1998**, *45*, 1124–1129.