

# Gauss-Benford Rules and Their Harmonic Peers

**Julio Rives**

Department of Artificial Intelligence, Universidad Nacional de Educacion a Distancia (UNED), Madrid, Spain

**Email address:**

[jrives@dia.uned.es](mailto:jrives@dia.uned.es)

**To cite this article:**

Julio Rives. (2024). Gauss-Benford Rules and Their Harmonic Peers. *Mathematics Letters*, 10(3), 24-35.

<https://doi.org/10.11648/j.ml.20241003.11>

**Received:** 5 November 2024; **Accepted:** 25 November 2024; **Published:** 18 December, 2024

---

**Abstract:** We research the relationship between the probability functions of the twofold hyperbolic universe, consisting of the logarithmic (real) and harmonic (rational) worlds. Within the logarithmic realm, we study the connection between the Gauss-Kuzmin distribution and Newcomb-Benford law and prove that they are fundamentally equivalent; the former corresponds to the probability decrements of the latter, i.e.,  $\log(2, 1 + 1/(k(k+2)))$  is the difference between the function  $\log(2, 1 + 1/n)$  evaluated at  $n=k$  and  $n=k+1$ , where  $k$  is the index of a coefficient of a real number's regular continued fraction expansion and  $n$  is a positive numeral written in positional notation. Thus, the binary Newcomb-Benford probability of  $n=1$  is the sum of all the Gauss-Kuzmin masses, of  $n=2$ , is the sum of all the Gauss-Kuzmin masses minus the Gauss-Kuzmin mass at  $k=1$ , and of  $n=m$ , is the sum of all the Gauss-Kuzmin masses minus the sum of the first  $m-1$  ones. Besides, the extrapolation of the Gauss-Kuzmin measure outside the unit interval subsumes Newcomb-Benford's cumulative distribution function. These findings lead to the Gauss-Benford law, which specifies that the occurrence possibility of a positive real number represented in positional notation is  $\log(2, 1+x)$  (i.e., the Gauss-Kuzmin measure) if  $x$  is within the unit interval and  $\log(2, 1+1/x)$  if  $x$  is outside. Geometrically, these possibilities indicate proximity to 1. The map between both domain partitions is a sheer inversion, the unique conformal transformation that fixes one and respects the minimum information principle. Moreover, we introduce the Gauss-Benford measure,  $\log(1+1/x, 1+y)$ , as the probability of a random variable accumulated between  $x-1$  and  $x$ , with density  $1/((1+y)\ln(1+1/x))$ , where  $0 \leq x-1 < y \leq x$ . We also explore the analogous harmonic rules; the canonical (normalized) harmonic probability of a positive natural  $q$  is  $1/q$ , the  $q$ th harmonic gap has probability mass  $1/(q(q+1))$ , and the harmonic occurrence possibility of a positive rational number is  $t$  if  $t$  is in the unit interval and  $1/t$  if  $t$  is outside. We build the bridge between the logarithmic and harmonic realms by integrating and normalizing the latter.

**Keywords:** Gauss-Kuzmin Distribution, Regular Continued Fraction, Positional Notation, Newcomb-Benford Law, Gauss-Benford Law, Smallness, Harmonic Scale, Logarithmic Scale

---

## 1. Introduction

This article focuses on the relationship between the Gauss-Kuzmin Distribution (GKD) and Newcomb-Benford Law (NBL); both are equivalent. We arrive at the GKD by taking decrements of the binary NBL masses. Analyzing the Gauss-Kuzmin measure (GKM) and NBL's cumulative distribution function, we derive a joint Gauss-Benford law (GBL) for the real numbers. Are logarithmic laws interconnected from the information theory perspective and in the real world?

NBL applies to many natural phenomena or observations saved as raw numerical data. The first digits of the numerals found in a series of records of the most varied sources exhibit a preference for minor values rather than displaying a uniform

distribution; the coefficients of a regular continued fraction (RCF) also hold this property. We seek to cover a gap in existing research, where NBL and GKD are studied separately, notwithstanding that both use the same Borel measure and are scale-invariant laws up to the radix. Indeed, we only knew a little about GKD before delving into the NBL. We have found that GKD complements NBL.

GKD is the discrete and direct probability mass function (PMF) arising as the limit probability distribution of the coefficients  $a_n \in \{1, 2, 3, \dots\}$  in the RCF expansion of a random (stochastic) variable [1] represented as

$$1/(a_1 + 1/(a_2 + 1/(a_3 + 1/(a_4 + \dots)))) \in (0, 1].$$

The random variable's domain is, thus, a generic real

number in the unit interval, and its range is the Borel (measurable) space given by the function

$$\lim_{n \rightarrow \infty} \Pr(a_n = k) = \log_2 \left( 1 + (k^2 + 2k)^{-1} \right) \in \mathbb{R}.$$

A classical path to GKD that we will elude hinges on the ergodicity of the Gauss map, a volume-preserving transformation applicable to deterministic dynamical stochastic systems that evolve as a whole [2]. Another course to this distribution is related to the RCF's projective invariance because the cross-ratio [3] and the connection with the modular group [4] lead us to the crucial concept of conformality [5]. Notwithstanding, our approach employs the recurrence relation concerning the fractional part of the RCF's remainders ([6], section 14.3).

The RCF representation of a real number is essential to analyze its degree of irrationality [7]. However, we will deal in this article with neither the accuracy of the representation nor how the partial expansions (convergents) approach the target [8, 9]. Likewise, given that GKD's entropy is approximately 3.43253 bits [10], we can think of a random variable's RCF expansion as a source of information [11] where a coefficient carries nearly 3.5 bits of information on average about the number. Interestingly, these coefficients exhibit a nonuniform distribution and some statistical dependence on close neighbors, resembling some properties of the NBL.

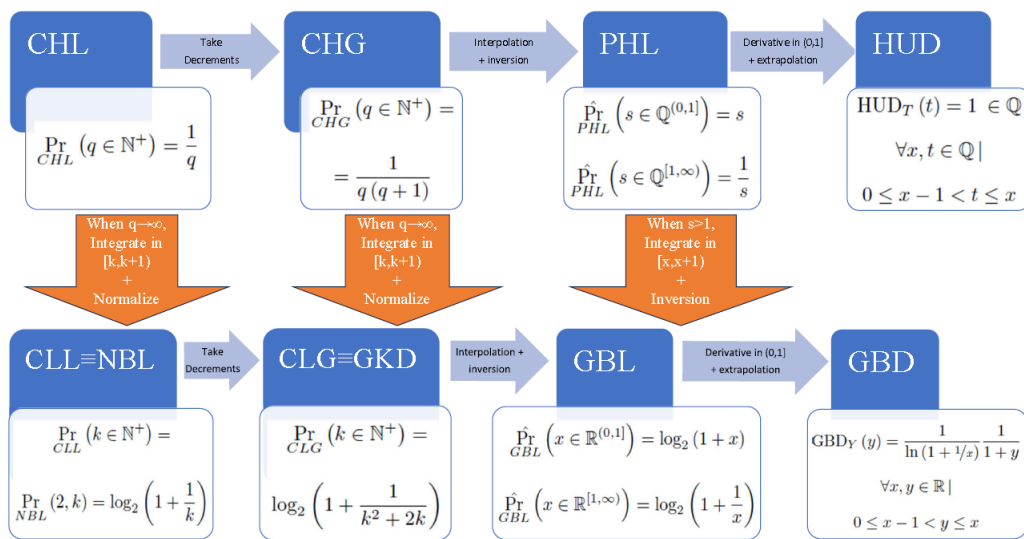
Putting number and information theories aside, the GKD is vital for understanding the statistical properties of the entries and patterns in RCF and the link with functional analysis. "The point is that the Gauss map, which is essentially a shift map on entries in continued fraction expansions, does not have Lebesgue measure as a measure-preserving transformation, but the measure  $\log_2(1+x)$  is a m.p.t. for the Gauss map" [12]. This measure leads to metrics in Diophantine approximation utilizing ergodic theory [13].

Generally, "If you can solve some problem using continued fraction then (via Gauss-Kuzmin statistics) you can study this problem from statistical point of view". In this sense, the principal application is factually to catch the statistics of various versions of the Euclidean algorithm [14], especially the average running time [15]. Nonetheless, we will not tackle applications of GKD in this article other than those related to NBL.

GKD is transversal to NBL. More precisely, GKD is the inverse function of NBL; we get a nonzero natural number from a range of probabilities. Working on the gaps between NBL probability masses of consecutive numerals, we have discovered that  $\Pr(a_n = k) \propto [\log(1 + 1/x)]_{k+1}^k$ . Normalization regarding the maximum NBL probability mass produces the  $1/\ln 2$  factor.

Since NBL explains GKD (theorem 3.1), positional notation (PN) plays a role in the RCF representation despite being radixless. Suppose we divide  $x$  by  $10^n$ , with  $n \in \mathbb{N}$ ; it is not essentially the same number? This view is implicit in Lochs' theorem [16], which relates the efficiency of an RCF expansion with that of a decimal notation in representing a real number. Thus, the representation of the random variables we deal with is ultimately logarithmic.

The binary NBL interpolated between numerals is a real function that, applied on the variable's reciprocal, is precisely the GKM. The concatenation of both measures leads us to the GBL, a piecewise possibility distribution function (theorems 4.1 and 4.3) with two partitions, the mirror of each other via a sheer inversion (theorem 4.2). These possibilities estimate to what extent a real number is in 1's neighborhood. Moreover, the GKD also admits interpolation between natural numbers. Furthermore, the GKM extrapolated outside the unit interval subsumes Benford's cumulative distribution function, suggesting that we can generalize the GKM to any real interval.



**Figure 1.** Map of the harmonic (rational) and logarithmic (Gauss-Benford) rules. The harmonic row displays the CHL (canonical harmonic law), CHG (canonical harmonic gaps), PHL (proportion-hyperbola law), and HUD (harmonic uniform density). The logarithmic row displays the CLL (canonical logarithmic law) or NBL in binary, CLG (canonical logarithmic gaps), or GKD, GBL, and GBD. The arrows between the harmonic and logarithmic formulas indicate how to bridge between them (see section 6).

We assume no knowledge about the subject and address every research objective in a different chapter. We analyze GKD and GKM, prove that Gauss-Kuzmin masses follow from the logarithmic (fiducial) NBL, derive the possibility distribution of a real number, and define the Gauss-Benford measure (GBM, definition 5.1) and the Gauss-Benford density (GBD, definition 5.2). From our unifying perspective, considering that the harmonic and logarithmic worlds run in parallel, we have further deduced Gauss-Benford's analogous harmonic rules (see Figure 1). For instance, the uniform

probability distribution (definition 6.4) is the GBD's rational peer. We discuss less tangible concepts in the concluding remarks.

## 2. The Gauss-Kuzmin Measure and Distribution

The RCF representation of a nonzero real number  $x$  is

$$x \equiv a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}} \equiv [a_0; a_1, a_2, a_3, \dots, a_n, \dots] \in \mathbb{R}$$

where  $a_0 \in \mathbb{Z}$  and  $a_n \in \mathbb{N}^+ \in \mathbb{N}^+$ . Mind that this expansion is finite for a rational number and infinite for an irrational.

Let  $\xi_0 = x$  be the initial remainder and

$$\xi_n = \frac{1}{\xi_{n-1} - a_{n-1}}$$

Then,

$$a_0 = \lfloor \xi_0 \rfloor = \lfloor x \rfloor$$

$$a_1 = \lfloor \xi_1 \rfloor = \left\lfloor \frac{1}{\xi_0 - a_0} \right\rfloor = \left\lfloor \frac{1}{x - \lfloor x \rfloor} \right\rfloor$$

$$a_2 = \lfloor \xi_2 \rfloor = \left\lfloor \frac{1}{\xi_1 - a_1} \right\rfloor = \left\lfloor \frac{1}{\xi_1 - \lfloor \xi_1 \rfloor} \right\rfloor = \left\lfloor \frac{1}{\frac{1}{x - \lfloor x \rfloor} - \left\lfloor \frac{1}{x - \lfloor x \rfloor} \right\rfloor} \right\rfloor$$

and generally

$$a_n = \lfloor \xi_n \rfloor = \left\lfloor \frac{1}{\xi_{n-1} - a_{n-1}} \right\rfloor = \left\lfloor \frac{1}{\xi_{n-1} - \lfloor \xi_{n-1} \rfloor} \right\rfloor$$

where  $\lfloor \cdot \rfloor$  is the floor function.

Let us circumscribe  $x$  to the unit interval to simplify our description so that  $a_0 = \lfloor \xi_0 \rfloor = 0$  and  $x \equiv [a_1, a_2, a_3, \dots, a_n, \dots] \in (0, 1]$ . Let  $\tau^0(x) = x = \xi_0 - a_0 = \xi_0 - \lfloor \xi_0 \rfloor$  be the initial fractional part and  $\tau^n(x) = \xi_n - \lfloor \xi_n \rfloor$  the  $n$ th fractional part.

Gauss noticed that  $a_n = 1$  many times and that the coefficients of a generic real number are small overall, so  $\Pr(\tau^n(x) < x) > x$ , meaning neither the random variable's cumulated probability is a sheer proportion nor its probability density function the uniform distribution. He asserted, even though he never unveiled the proof, that

$$\lim_{n \rightarrow \infty} \Pr(\tau^n(x) < \tau^0(x)) = \lim_{n \rightarrow \infty} \Pr(\xi_n - \lfloor \xi_n \rfloor < x) = \log_2(1+x) \quad (1)$$

This logarithmic cumulative distribution, known as GKM, has a middle point (given by  $\log_2(1+x_{1/2}) = 1/2$ ) at  $x_{1/2} = \sqrt{2} - 1$ . This fundamental asymmetry of the real unit interval leads to a Pareto rule  $(\sqrt{2}-1)/(2-\sqrt{2}) = 1/\sqrt{2} \approx 29\%/71\%$ . Equivalently, the odds of getting a fractional part in the first half of the unit interval against the second half are  $\log_2(1+1/2)/(1-\log_2(1+1/2)) \approx 1.41$ .

Next, we summarize Havil's derivation of deducing GKD from (1).  $\forall k \in \mathbb{N}^+$ , it holds

$$\begin{aligned} \Pr(a_n = k) &= \Pr(k < \xi_n < k+1) \\ &= \Pr(\xi_n < k+1) - \Pr(\xi_n < k) \\ &= \Pr(1/\xi_n > 1/(k+1)) - \Pr(1/\xi_n > 1/k) \\ &= (1 - \Pr(1/\xi_n < 1/(k+1))) - (1 - \Pr(1/\xi_n < 1/k)) \\ &= \Pr(1/\xi_n < 1/k) - \Pr(1/\xi_n < 1/(k+1)) \\ &= \Pr(\xi_{n-1} - \lfloor \xi_{n-1} \rfloor < \frac{1}{k}) - \Pr(\xi_{n-1} - \lfloor \xi_{n-1} \rfloor < \frac{1}{k+1}) \end{aligned} \quad (2)$$

When  $n \rightarrow \infty$ , the RCF expansion approaches the target real number, and using (1), we get (see Figure 2)

$$\begin{aligned} \Pr_{GKD}(a_n = k) &\rightarrow \log_2 \left(1 + \frac{1}{k}\right) - \log_2 \left(1 + \frac{1}{k+1}\right) \\ &= \log_2 \left(1 + \frac{1}{k(k+2)}\right) \end{aligned} \quad (3)$$

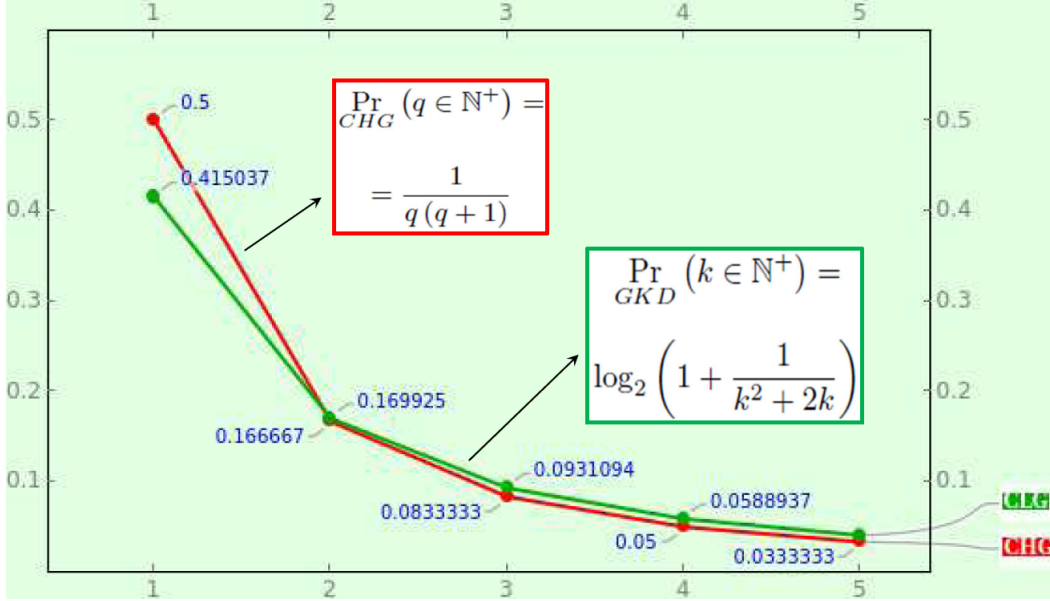


Figure 2. The canonical harmonic (rational) and logarithmic (Gauss-Kuzmin) PMFs.

The probabilities sum to 1, the arithmetic mean diverges, the mode is  $k = 1$ , and the median is  $k = 2$ . The geometric and harmonic means of the GKD series do stick to the constants  $\kappa_0 = 2.685452001\dots$  (aka Khinchin's) and  $\kappa_{-1} = 1.745405662\dots$ , respectively. These constants imply that the RCF coefficients of a random variable are not entirely spontaneous. In principle, adherence to GKD (and  $\kappa_0$ ) measures the extent to which a real number is nontrivial and incidental. Neither rationals nor roots of quadratic equations (e.g., the golden ratio  $\phi$ ) comply with GKD.  $e$ 's RCF coefficients present a pattern with no geometric mean. In contrast, the coefficients of  $\kappa_0$  itself,  $\pi$ , Euler-Mascheroni constant, Meissel-Mertens constant,  $\sin 1$ ,  $\ln 2$ ,  $\ln 7$ ,  $\ln_3 1234567$ ,  $\sqrt[3]{2}$ ,  $\sqrt[3]{3}$ , Tribonacci constant, magic

angle, Riemann zeta function at 2 and 3, Dirichlet beta function at 2 and 3, Catalan's constant, Dottie number, polygon inscribing constant, Soldner constant, Gelfond-Schneider constant, second Favard constant, Erds-Borwein constant, universal parabolic constant, Gompertz constant, Artin's constant, et cetera are seemingly free of patterns but, based on numerical evidence, approximate to GKD and their geometric means converge to  $\kappa_0$ .

Another proof of how to arrive from (1) at (3) uses ergodic theory. Effectively, an alternative manner of posing the problem is taking an RCF expansion as a trajectory of a one-dimensional dynamical system via the Gauss transformation  $\tau : (0, 1] \mapsto (0, 1]$  iterating over the remainder's fractional part. Given  $\tau(x) := \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ , the trajectory

$$\{\tau^0(x) = x, \tau^1(x) = \tau(x), \tau^2(x) = \tau(\tau(x)), \dots, \tau^n(x) = \tau(\tau^{n-1}(x)), \dots\}$$

is such that  $\lim_{n \rightarrow \infty} \Pr(\tau^n(x) = \xi_n - \lfloor \xi_n \rfloor < \tau^0(x)) = \log_2(1+x)$ . This measure's distribution function is the leading eigenfunction for the transfer operator associated with the Gauss map [17].

### 3. GKD from NBL

For (3) is proportional to  $[\log(1 + \frac{1}{x})]_{k+1}^k$ , we can expect a profound link between GKD and NBL.

A sample of numerals expressed in standard PN satisfies the fiducial NBL for radix  $r$  if the leading digit  $d$  occurs with

probability

$$\Pr_{NBL}(r, d) = \log_r \left(1 + \frac{1}{d}\right) \in \mathbb{R} \quad (1 \leq d < r) \quad (4)$$

where  $\{d, r \in \mathbb{N}^+\}$ . We can generalize this PMF to the probability of getting a leading  $r$ -ary numeral  $k \in \mathbb{N}^+$  of any length as the difference

$$\begin{aligned}\Pr_{NBL}(r, k) &= \log_r(k+1) - \log_r k \\ &= \log_r \left(1 + \frac{1}{k}\right) \in \mathbb{R}\end{aligned}\quad (5)$$

For example, a ternary numeral's probability of starting with "21" (5 in standard decimal), say  $.212_3$  or  $2101_3$ , is  $\log_3 \left(1 + \frac{1}{5}\right) \approx 16.6\%$ . Equation (5) reduces to (4) when  $k$  is an  $r$ 's digit.

**Theorem 3.1.** The masses of GKD are the mass decrements of NBL with radix 2.

*Proof* The series  $\xi_n$  consists of  $r$ -ary real numbers in  $[1, \infty)$ . Observe that

$$\Pr(k \leq \xi_n < k+1) \propto \Delta_{NBL}(r, k)$$

where

$$\sum_{k=1}^{\infty} \frac{\Delta_{NBL}(r, k)}{\Pr_{NBL}(r, 1)} = \sum_{k=1}^{\infty} \frac{\log_r \left(1 + \frac{1}{k^2+2k}\right)}{\log_r 2} = \sum_{k=1}^{\infty} \log_2 \left(1 + \frac{1}{k^2+2k}\right) = 1$$

PMF  $\Pr(a_n = k) = \Pr(k \leq \xi_n < k+1) = \Delta_{NBL}(r, k)/\Pr_{NBL}(r, 1)$  is precisely GKD.  $\square$

$$\begin{aligned}\Delta_{NBL}(r, k) &= \Pr_{NBL}(r, k) - \Pr_{NBL}(r, k+1) \\ &= \log_r \left(1 + \frac{1}{k}\right) - \log_r \left(1 + \frac{1}{k+1}\right) \\ &= \log_r \left(\frac{1+\frac{1}{k}}{1+\frac{1}{k+1}}\right) \\ &= \log_r \left(\frac{k+1}{k+2}\right) \\ &= \log_r \left(\frac{(k+1)^2}{k(k+2)}\right) \\ &= \log_r \left(\frac{k^2+2k+1}{k^2+2k}\right) \\ &= \log_r \left(1 + \frac{1}{k^2+2k}\right)\end{aligned}$$

When  $k$  climbs to  $\infty$ , this probability gap vanishes using any radix. Because of  $\Pr(r, k) \in [0, 1]$ , the infinite series  $\Delta_{NBL}(r, k)$  with  $k = \{1, 2, 3, \dots\}$  becomes another PMF regarding the maximum NBL probability mass,  $\Pr_{NBL}(r, 1) = \log_r 2$ . Since

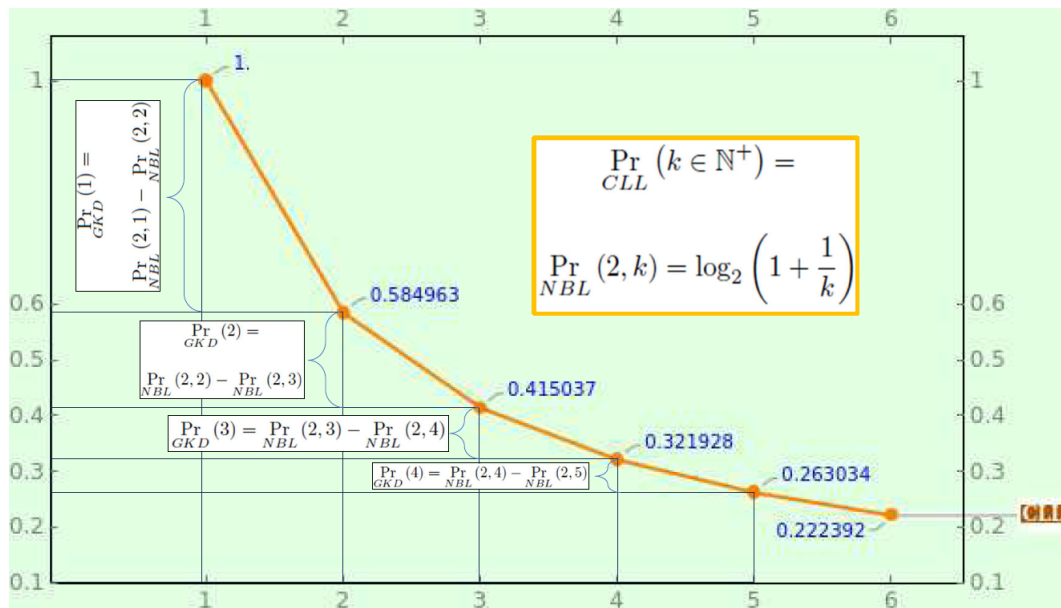


Figure 3. Transversality between NBL and GKD.

Thus, we can rename the GKD as CLG (see Figure 3).

**Definition 3.1.** A natural number  $k$  is said to satisfy the CLG if it occurs with the probability

$$\Pr_{CLG}(k) = \frac{\Delta_{NBL}(r, k)}{\Pr_{NBL}(r, 1)} = \log_2 \left(1 + \frac{1}{k^2+2k}\right) \in \mathbb{R} \quad (k \in \mathbb{N}^+)$$

Taking this PMF as input and following (2)'s derivation oppositely leads us to (1). In Havil's words ([6], page 157), although it is "hard to imagine", this could be the "mysterious thought process" Gauss developed to arrive at

GKM, anticipating Simon Newcomb's note written 69 years later.

## 4. Gauss-Benford Law

We have presented GKM (1) as a cumulative distribution function. We can also contemplate the GKM as the possibility that a random variable taking on the value  $m \in (0, 1]$  is in 1's neighborhood. This angle has the advantage of assigning a straightforward estimate to a real number rather than the result of an accretion process and suggesting a geometry perspective; GKM measures the logarithmic closeness from  $m$  to 1. Equivalently, GKM defines the membership degree of  $m$  to the concept of "1's neighbour" in terms of fuzzy sets. More remarkably, these interpretations are likewise valid when we move outside the unit interval.

**Theorem 4.1.** GKM is a possibility distribution function in the unit interval.

*Proof* Given the set of real numbers in the positive unit interval and a nonempty family  $I$  of subintervals of  $(0, 1]$ , i.e., a subset of the powerset  $P((0, 1])$ , the function  $\hat{\mu}(I) = \sup_{m \in I} (\log_2(1 + m))$  is a fuzzy measure on  $\langle(0, 1], I\rangle$  because it satisfies ([18], definition 7.1) the boundary conditions ( $\hat{\mu}(\emptyset) = 0$  and  $\hat{\mu}((0, 1]) = 1$ ), monotonicity ( $\forall A, B \in I$ , if  $A \subseteq B$ , then  $\hat{\mu}(A) \leq \hat{\mu}(B)$ ), continuity from below (for any increasing sequence  $A_1 \subset A_2 \subset \dots$  in  $I$ , if  $\bigcup_{i=1}^{i=\infty} A_i \in I$ , then

$\lim_{i \rightarrow \infty} \hat{\mu}(A_i) = \hat{\mu}\left(\bigcup_{i=1}^{i=\infty} A_i\right)$ ), and continuity from above (for

any decreasing sequence  $A_1 \supset A_2 \supset \dots$  in  $I$ , if  $\bigcap_{i=1}^{i=\infty} A_i \in I$ ,

then  $\lim_{i \rightarrow \infty} \hat{\mu}(A_i) = \hat{\mu}\left(\bigcap_{i=1}^{i=\infty} A_i\right)$ ).

Besides, any family  $\{A_k | k \in K\}$  in  $I$  such that  $\bigcup_{k \in K} A_k \in I$ , where  $K$  is an arbitrary index set, holds

$$\begin{aligned} \hat{\mu}\left(\bigcup_{k \in K} A_k\right) &= \sup_{m \in \bigcup_{k \in K} A_k} (\log_2(1 + m)) \\ &= \log_2\left(1 + \sup_{k \in K} m \in \bigcup_{k \in K} A_k\right) \\ &= \sup_{k \in K} \hat{\mu}(A_k) \end{aligned}$$

meaning  $\hat{\mu}(I)$  is a possibility measure ([18], 7.22).

Moreover, because  $\log_2(1 + m)$  uniquely determines  $\hat{\mu}(A) \forall A \in P((0, 1])$  ([18], 7.29), GKM is a possibility distribution function.  $\square$

An inversion conformally maps the domain  $(0, 1]$  onto the domain  $[1, \infty)$ .

**Theorem 4.2.** A sheer inversion is the only conformal map that, assuming the minimum information principle, turns the unit interval inside out and fixes 1.

*Proof* The inversion  $m \mapsto 1/m$  fixes the integer one and interchanges the inside and outside of the unit interval. In particular, it swaps the origin and infinity. This map is conformal because it conserves the cross-ratio [19]. Although other inversions, such as  $z \mapsto \alpha e^\beta / z$ , can also serve

as bijections between both domains, preserving relative distances, they do not yield a better result and need extra parameters we cannot explain (for instance,  $\alpha \neq 1$  and  $\beta \neq 0$ ). A mere inversion is the unique conformal map that agrees with the principle of minimum information [20], i.e., avoids introducing unwarranted assumptions.

Consider the hyperbolic closeness of  $m$  to 1 plus the hyperbolic closeness of 1 to  $-m$ . On the one hand,

$$(-m, 0; 1, \infty) \xrightarrow{m \mapsto 1/m} (-1/m, \infty; 1, 0)$$

where

$$(-m, 0; 1, \infty) = 1 + m = \frac{1 + 1/m}{0 + 1/m} = (\infty, 1/m; 0, 1)$$

On the other hand,

$$(0, m; 1, \infty) \xrightarrow{m \mapsto 1/m} (\infty, 1/m; 1, 0)$$

where

$$(0, m; 1, \infty) = \frac{1 - 0}{1 - m} = \frac{0 - 1/m}{1 - 1/m} = (\infty, 1/m; 1, 0)$$

Then,

$$\log_2(-m, 0; 1, \infty) + \log_2(0, m; 1, \infty) = \log_2\left(\frac{1 + m}{1 - m}\right)$$

is the binary hyperbolic distance from the origin to  $m$  in agreement with the conformal one-dimensional disk (unit-interval, or 1-ball) model [21]. Outside, the distance from  $1/m$  to infinity (as the reciprocal of the origin) is the same. So, the function  $\log_2(1 + 1/m)$  outside our conformal model space mirrors the GKM inside.  $\square$

Because the inversion  $m \mapsto 1/m$  is an orientation-reversing conformal mapping, we can also interpret the GKM of  $m$ 's reciprocal as the possibility that a random variable taking on the value  $m \in \mathbb{R}^{[1, \infty)}$  is in 1's neighborhood.

**Theorem 4.3.**  $\log_2(1 + 1/m)$  is a possibility distribution function in  $\mathbb{R}^{[1, \infty)}$ .

*Proof* Given the set of real numbers in  $[1, \infty)$  and a nonempty family  $I$  of subintervals, i.e., a subset of  $[1, \infty)$ 's powerset  $P([1, \infty))$ , the function  $\hat{\mu}(I) = \sup_{m \in I} (\log_2(1 + 1/m))$  is a fuzzy measure on  $\langle[1, \infty), I\rangle$  because it satisfies the boundary conditions, monotonicity, and continuity from below and above.

Besides, any family  $\{A_k | k \in K\}$  in  $I$  such that  $\bigcup_{k \in K} A_k \in I$ , where  $K$  is an arbitrary index set, holds

$$\begin{aligned}
\hat{\mu} \left( \bigcup_{k \in K} A_k \right) &= \sup_{m \in \bigcup_{k \in K} A_k} (\log_2 (1 + 1/m)) \\
&= \log_2 \left( 1 + \sup_{k \in K} \frac{1}{m \in \bigcup_{k \in K} A_k} \right) \\
&= \sup_{k \in K} \hat{\mu} (A_k)
\end{aligned}$$

meaning  $\hat{\mu}(I)$  is a possibility measure.

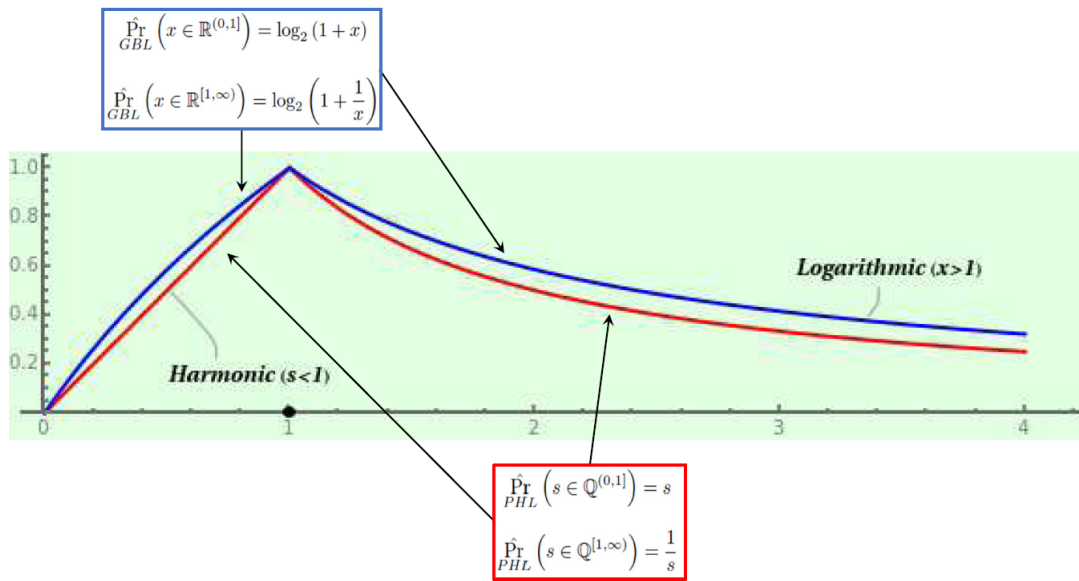
Further, because  $\log_2(1 + 1/m)$  uniquely determines

$\hat{\mu}(A) \quad \forall A \in P([1, \infty))$ , the inversion of GKM's variable yields a possibility distribution function in  $\mathbb{R}^{[1, \infty)}$ .  $\square$

Interestingly, owing to its properties, a possibility distribution function may be interpreted as a distribution of upper probability ( $\hat{\text{Pr}}$ ) estimates ([18], page 205).

**Definition 4.1.** A random variable  $X$  satisfies the GBL if it occurs with the possibility (see Figure 4)

$$\hat{\text{Pr}}_{GBL}(X = x) = \begin{cases} \log_2(1 + |x|) & (0 < |x| \leq 1) \\ \log_2\left(1 + \frac{1}{|x|}\right) & (1 \leq |x| < \infty) \end{cases} \in \mathbb{R}$$



**Figure 4.** The rational numbers' harmonic law and the real numbers' GBL. We can geometrically interpret these possibility distribution function as "proximity to 1".

If  $x$  is a nonzero natural number,  $\hat{\text{Pr}}_{GBL}(x)$  is precisely the CLL.

The inversion map does not affect any remainder  $\xi_n$ . The reciprocal of the real number  $[a_0; a_1, a_2, a_3, \dots]$  with  $a_0 \geq 1$  is  $[a_0, a_1, a_2, a_3, \dots]$ , sharing the RCF expansion

and, hence, the same occurrence possibility. However, GBL allows us to generalize the probability interval constraint of the remainders because it interpolates between NBL masses; the limit probability of a remainder falling in an arbitrary unit-length real interval  $[x, x+1)$  ( $x \geq 1$ ) is

$$\lim_{n \rightarrow \infty} \text{Pr}(x \leq \xi_n < x+1) = \hat{\text{Pr}}_{GBL}(x) - \hat{\text{Pr}}_{GBL}(x+1) = \log_2 \left( 1 + \frac{1}{x(x+2)} \right) \quad (6)$$

meaning GKD also admits interpolation between natural numbers. In other words, (6) is the GKD extended to the positive real numbers.

## 5. Gauss-Benford Measure and Density

GKM is virtually the same as PMF (4)'s cumulative distribution function,  $\text{Pr}_{NBL}(d \leq x, r) = \log_r(1 + \lfloor x \rfloor)$ ,

where  $d$  is a natural number in  $[1..r)$ ,  $x$  is a real number in  $[1, r)$ ,  $r$  is the radix, and  $\lfloor \cdot \rfloor$  is the floor function.

Now, let us forget about the radix's digits. Equation (5) is valid for any radix and numeral. The peculiarity of the binary logarithm is that the probability of a numeral  $k$  is automatically normalized concerning "1" because  $\text{Pr}_{NBL}(2, 1) = \log_2(1 + \frac{1}{1}) = 1$ . The cumulative NBL probability from numeral  $k = 1$  to  $k = m$  in binary corresponds to the  $m$ th partial sum

$$S_2(m) = \sum_{k=1}^{k=m} \text{Pr}_{NBL}(2, k) = \sum_{k=1}^{k=m} \log_2 \left( 1 + \frac{1}{k} \right) = \log_2(1 + m)$$



Thus, we might extrapolate the range of GKM values outside the unit interval. Regarding an RCF's expansion, such a generalized GKM gives place to the following generalized contour conditions,

$$\lim_{\substack{n \rightarrow \infty \\ s \rightarrow \lfloor s \rfloor}} \Pr(\tau^n(s - \lfloor s \rfloor) < \tau^0(s - \lfloor s \rfloor)) = \log_2(1 + \lfloor s \rfloor)$$

$$\lim_{\substack{n \rightarrow \infty \\ s \rightarrow \lfloor s \rfloor + 1}} \Pr(\tau^n(s - \lfloor s \rfloor) < \tau^0(s - \lfloor s \rfloor)) = \log_2(2 + \lfloor s \rfloor)$$

For example, when  $s \rightarrow 15^+$ , the generalized GKM tends to  $\log_2(1 + 15) = 4$ , whereas when  $s \rightarrow 16^-$ , the extrapolated GKM tends to  $\log_2(17) = 4.08746\dots$ . The  $k$ th ( $k \in \mathbb{N}^+$ ) unit interval accumulates the probability  $[\log_2(1 + s)]_{k-1}^k$ . Furthermore, every unit interval on the positive real line accumulates the probability  $[\log_2(1 + s)]_{x-1}^x = \log_2(1 + 1/x)$ . Therefore, the GKM's evaluation at the endpoints of an arbitrary unit-length real interval precisely yields the interpolation of NBL between natural numbers.

However, we must renormalize the GKM outside the unit interval because  $\log_2(1 + 1/x)$  does not fulfill countable additivity in  $[1, \infty)$ .

**Definition 5.1.** The GBM for a random variable  $Y$  in an arbitrary positive interval  $(x - 1, x]$  ( $x \geq 1$ ) is the function

$$\hat{\Pr}_{GBM}(Y = y) = \log_{1+1/x}(1 + y)$$

Its derivative regarding  $y$  is the density function GBD (see Figure 5).

**Definition 5.2.** A random variable  $Y$  has GBD if its density is

$$\text{GBD}_Y(y) = \frac{1}{\ln(1 + 1/x)} \frac{1}{1 + y} \quad \forall x, y \in \mathbb{R} \mid 0 \leq x - 1 < y \leq x$$

Specifically, GBD is in  $(0, 1]$  the density of the measure invariant under the Gauss map and the Gauss-Kuzmin-Wirsing operator's fixed point up to scaling [22]. Note that the further the interval is from the origin, the closer GBD is to the continuous uniform distribution on that interval. GBD fulfills countable additivity because of

$$\int_{x-1}^x \text{GBD}(y) dy = \frac{\int_{x-1}^x \frac{dy}{1+y}}{\ln(1 + 1/x)} = \frac{[\ln(1 + y)]_{x-1}^x}{\ln(1 + 1/x)} = 1$$

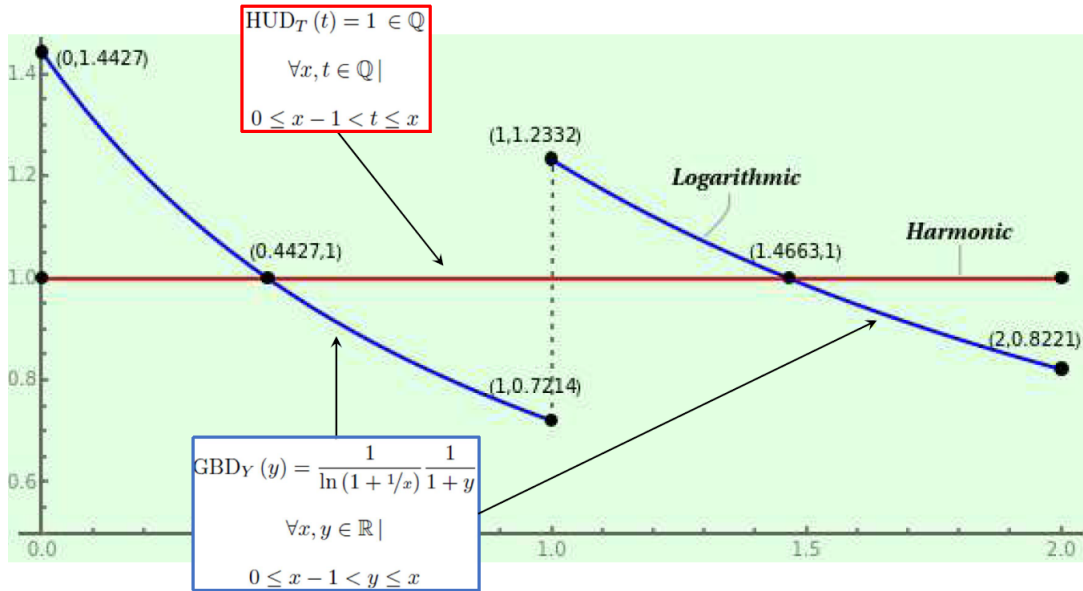


Figure 5. The harmonic (rational) and logarithmic (Gauss-Benford) probability density functions.

## 6. Gauss-Benford's Harmonic Peers

According to [23], harmonic versions of the Gauss-Kuzmin laws must exist, i.e., rational analogues with equations (3) and

(5), and definitions (4.1) and (5.2).

Definition (3.1) proves that GKD is the PMF of the NBL's differences. Eq. (5)'s harmonic counterpart is PMF  $1/(qH_{b-1})$ ,



where  $H_n$  is the  $n$ th harmonic number,  $\{q, b\} \in \mathbb{N}^+$ , and  $1 < b$ . A quantum  $q$  is the simplest computable entity. Quantum probability masses are self-normalized with  $b = 2$ ; the resulting law is CHL, outlining a unit square hyperbola that represents the occurrence probability of a quantum.

**Definition 6.1.** A PN-encoded quantum  $q$  is said to satisfy the CHL if it occurs with probability

$$\Pr_{CHL}(q) = \frac{1}{q} \in \mathbb{Q} \quad (q \in \mathbb{N}^+)$$

**Definition 6.2.** A natural number  $q$  is said to satisfy the CHG if it occurs with probability

$$\Pr_{CHG}(q) = \Pr_{CHL}(q) - \Pr_{CHL}(q+1) = \frac{1}{q} - \frac{1}{q+1} = \frac{1}{q(q+1)} \in \mathbb{Q} \quad (q \in \mathbb{N}^+)$$

This expression is the PMF for a random variable that takes on values from a sample space of natural numbers (see Figure 2). This simple univariate discrete distribution exhibits asymptotic convergence and is well-formed because of

$$\sum_{N=1}^{\infty} \frac{1}{q^2 + q} = 1$$

The mean and variance diverge, the mode is 1, and the entropy is

$$-\sum_{q=1}^{\infty} \Pr_{CHG}(q) \ln \Pr_{CHG}(q) = \sum_{q=1}^{\infty} \frac{\ln(q^2 + q)}{q^2 + q} \approx 1.98356 \text{ nats} \approx 2.86167 \text{ bits}$$

The point of CHG is that it hides under its tail PMF CLG (definition 3.1). When  $b \rightarrow \infty$ , the summation over the quanta becomes an integral that, taken between the endpoints of the  $k$ th unit interval, results in the GKD up to normalization, i.e.,

$$\int_k^{k+1} \frac{dq}{q(q+1)} = \log \left( 1 + \frac{1}{k^2 + 2k} \right) \quad (k \in \mathbb{N}^+)$$

Thus, we have found an alternative way to obtain the GKD.

How does definition (6.2) fit in Havil's rationale to derive GKD? GKM's harmonic analogous comes up if the remainder's fractional parts accumulate proportionally to the remainder's itself, i.e.,

$$\lim_{n \rightarrow \infty} \Pr(\tau^n(q) < \tau^0(q)) = \lim_{n \rightarrow \infty} \Pr(\xi_n - \lfloor \xi_n \rfloor < q) = q$$

Then, on the one hand,

$$\begin{aligned} \Pr(a_n = q) &= \Pr(q < \xi_n < q+1) \\ &\stackrel{*}{=} \Pr\left(\xi_{n-1} - \lfloor \xi_{n-1} \rfloor < \frac{1}{q}\right) - \Pr\left(\xi_{n-1} - \lfloor \xi_{n-1} \rfloor < \frac{1}{q+1}\right) \end{aligned}$$

where we obtain the equality with an asterisk like in (2). When  $n \rightarrow \infty$ , we get  $\Pr(a_n = q) \rightarrow \frac{1}{q} - \frac{1}{q+1} = \frac{1}{q(q+1)} = \Pr_{CHG}(q)$ .

On the other hand, a neat inversion allows us to interpolate between quanta.

**Definition 6.3.** A rational number  $s$  is said to satisfy the PHL if it occurs with the possibility (see Figure 4)

$$\hat{\Pr}_{PHL}(s) = \begin{cases} |s| & (0 < |s| \leq 1) \\ 1/|s| & (1 \leq |s| < \infty) \end{cases} \in \mathbb{Q}$$

If  $s$  is a nonzero natural number,  $\hat{\Pr}_{PHL}(s)$  is precisely the CHL.

We can extrapolate the PHL derivative in  $(0, 1]$  to an arbitrary rational unit-length interval (see Figure 5).

**Definition 6.4.** A rational random variable  $T$  has HUD if its density is

$$\text{HUD}_T(t) = 1 \quad \forall x, t \in \mathbb{Q} \mid 0 \leq x-1 < t \leq x$$

Because the PHL (definition 6.3) subsumes the interpolated CHL (definition 6.1), the limit probability of a remainder falling in an arbitrary rational unit-length interval in  $[1, \infty)$  is

When  $b$  approaches infinity, by integrating this PMF over the  $k$ th unit interval and normalizing, we arrive at (5). Since CHL is the harmonic counterpart of the CLL  $\Pr_{NBL}(2, k) = \log_2(1 + 1/k)$ , we should also obtain GKD by taking differences in CHL, integrating, and normalizing.

Taking the probability mass decrements of (6.1) results in CHG.

$$\lim_{n \rightarrow \infty} \Pr(t \leq \xi_n < t+1) = \hat{\Pr}_{PHL}(t) - \hat{\Pr}_{PHL}(t+1) = \frac{1}{t(t+1)} \quad (7)$$

meaning the CHG (definition 6.2) also admits an interpolation between natural numbers. In other words, (7) is the GHG extended to the positive rational numbers.

## 7. Concluding Remarks

Let us take stock of this codicil. We have introduced the GKD and proved that it complements the NBL to the point that we can find the GKD calculating the differences between binary NBL masses. Next, we have deduced the GBL, which defines the possibility of a real number represented in PN. We have figured the GKM generalized to an arbitrary real unit-length interval and the associated density. Finally, we have inferred the Gauss-Benford's harmonic peers.

GKD is fundamental because it tells us the probability mass of a counting number arising as a coefficient in the RCF expansion of a random variable. Although we have not defined a formal compliance test, an infinitude of mathematical constants obeys GKD to a lower or greater degree; they are typical numbers living in the logarithmic world. For example, Niven's constant is apparently (first 1000 terms) more GKD-compliant than Foias constant, whence more NBL-compliant. Some irrational numbers are notably less or not NBL-compliant at all; they disobey GKD because they are either mathematically encoded in a more primitive way (where the sequence of RCF coefficients reflects a recurrence relation, e.g., the golden ratio, quadratic roots such as  $\sqrt{2}$ ,  $\sqrt{3}$ ,  $\sqrt{33}$ ,  $\sqrt{333}$ ,  $\sqrt{3333}$ , the first continued fraction constant,  $\tanh(1)$ , and Cahen's constant) or special-purpose manufactured (e.g., 10-ary Champernowne constant).

We have seen that GKD is transversal to the NBL. NBL takes a counting number and outputs a probability percentage, while GKD gets a counting number from a range of probability masses. Indeed, we construct the GKD by taking the decrements of the binary logarithmic accumulated probability masses. Likewise, we assemble the rational version of GKD by calculating the gaps between harmonic probability masses and normalizing their sum. Notably, the PHL estimates the possibility of a PN-encoded rational number much as the GBL gauges the proximity to 1 of a positive PN-encoded real number.

The fiducial NBL and GKD are logarithmic and, consequently, scale- and radix-invariant. Both traverse and complete each other and have practically the same cumulative distribution function. Both enhance the fact that place-value numeration plays a role in the RCF representation of a number.

Why has nobody noticed the link between NBL and GKD, which is apparent to a degree? We can figure some reasons out.

First, NBL is often associated with digits, albeit generally valid for numerals, such as the coefficients of an RCF expansion. This notion is necessary to establish an association with the GKD. Besides, whereas the fiducial NBL involves the

symbol zero, this has no role in the GKD. Because we are acquainted with Benford's distribution for bijective notation, which avoids such a symbol, we did not have this handicap.

A noteworthy obstacle is that NBL always concerns a base or radix, while the RCF representation of a number is, at first sight, radix-independent. However, the essence of NBL is precisely radix invariance, and GKD takes like NBL input data encoded in PN. Further, GKD's binary logarithm is dissuasive if we do not realize that it corresponds to the (maximum) probability mass of the symbol "1" and does not indicate a coding radix.

Nonetheless, the testing compliance process is the major impediment to associating both distributions. While conformance with Benford requires entire empirical or real-life distributions (data tables of molecular weights, cosmology, genome sizes, orography, demography, prices, sales, socio-economic or macroeconomic data of a country, election data, et cetera), conformance with GKD utilizes as input an isolated mathematical constant or a computer-generated value. However, our theory disproves this conception.

We can also utilize uncooked numerical data of all kinds of sciences to determine whether they obey GKD. For instance, take the World DataBank's population statistics per nation corresponding to the year 2011 (see <https://www.journalofaccountancy.com/issues/2017/apr/excel-and-benford's-law-to-detect-fraud.html>); it approximately complies with NBL, because the first-digit occurrence percentages are

$$\{28, 18, 15, 9, 9, 6, 7, 6, 4\}$$

against the NBL-theoretical

$$\{30, 18, 12, 10, 8, 7, 6, 5, 5\}.$$

Then, divide by the power of 10 that maximizes the number of decimal positions ( $10^{10}$ , in this case, to make China's population, namely 1.344.130.000, a real between 0 and 1), calculate the RCF expansion until a prefixed number of coefficients (say 16) of every (rational) record (e.g., Spain expands as  $.0046742697 \approx [213, 1, 14, 1, 10, 1, 1, 1, 31, 2, 14, 1, 1, 50, 1, 27, \dots]$ ), aggregate the number of 1's, 2's, 3's et cetera regarding the total of coefficients (218 nations times 16 coefficients equals 3488), and test compliance against the whole set of RCF expansions; e.g., numerals 1 to 8 occur with percentages

$$\{39.2, 15, 8.5, 4.8, 3.6, 2.5, 2.6, 1.4\}$$

against the GKD-theoretical

$$\{41.5, 17, 9.3, 5.9, 4.1, 3, 2.3, 1.8\}.$$

Disregarding China's entry (it seems fabricated) and the first coefficients (e.g., Spain's 213, spurious due to the preprocessing), compliance with Gauss-Kuzmin is even more impressive.

We conclude with some philosophical implications. This article supports the thesis that ours is a thrifty logarithmic

universe where PN implements efficiency and enables conformality; a numeral represents its price, irrespective of the radix. The harmonic world also prefers the small because, for instance, a quantum's occurrence possibility is proportional to its reciprocal. These facts suggest that the logarithmic laws and their harmonic counterparts describe a natural tendency to access, transform, and transmit information at the lowest possible cost. Our research supports this idea by establishing a parallelism between the logarithmic and harmonic worlds (embodied by the connection between CLL and CHL, CLG and CHG, and GBL and PHL) that explains the pervasive propensity to 1 and cohabitation of the continuum and the discrete. More generally, one cannot help but surmise the prospect that the hyperbolic laws have a common genesis at a fundamental level.

## Abbreviations

CHG	Canonical Harmonic Gaps
CHL	Canonical Harmonic Law
CLG	Canonical Logarithmic Gaps
CLL	Canonical Logarithmic Law
GBD	Gauss-Benford Density
GBL	Gauss-Benford Law
GBM	Gauss-Benford Measure
GKD	Gauss-Kuzmin Density
GKM	Gauss-Kuzmin Measure
HUD	Harmonic Uniform Density
NBL	Newcomb-Benford Law
PHL	Proportion-Hyperbola Law
PMF	Probability Mass Function
PN	Positional Notation
RCF	Regular Continued Fraction

## ORCID

0000-0003-3980-5829 (Julio Rives)

## Conflicts of Interest

The author asserts that only scientific rigor, significance, and clarity drive the goals of this work. It contains no known minor or significant incongruencies, errors, or defects.

The author declares no conflicts of interest. In particular, the author did not receive support from any organization for the submitted work, having 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] Wikipedia contributors, the Free Encyclopedia. Random variable, 2024. [Online; accessed 15-November-2024].
- [2] Cesar E. Silva. What Is... an Ergodic Transformation. *The American Mathematical Society*, 13(1): 26-27, January 2016.
- [3] Oleg N. Karpenkov. Gauss-Kuzmin Statistics, chapter 13, pages 139-154. Springer Berlin Heidelberg, 2022.
- [4] Caroline Series. The Modular Surface and Continued Fractions. *Journal of the London Mathematical Society*, s2-31(1): 69-80, February 1985.
- [5] David E. Blair. Inversion Theory and Conformal Mapping. *Graduate Studies in Mathematics*. American Mathematical Society, 2000.
- [6] Julian Havil. *Gamma - Exploring Euler's Constant*. Princeton University Press, 2003.
- [7] Burkard Polster and Marty Ross (Mathologer). Infinite Fractions and the Most Irrational Number, July 2016.
- [8] Eduard Wirsing. On the Theorem of Gauss-Kusmin-Lvy and a Frobenius-type Theorem for Function Spaces. *Acta Arithmetica*, 24(5): 507-528, 1974.
- [9] Michelle He, Brian Huang, Aadeesh Shastry, and Mirilla Zhu. Benford's Law in Linear Recurrences and Continued Fractions. University of Washington, August 2017.
- [10] Linas Vepstas. Entropy of Continued Fractions (Gauss-Kuzmin Entropy). Unpublished, 2014.
- [11] Nelson M. Blachman. The Continued Fraction as an Information Source (Corresp.). *IEEE Transactions on Information Theory*, 30(4): 671-674, July 1984.
- [12] Dan L. (Mathoverflow). Gauss-Kuzmin Theorem (continued fractions) - Why Is important? MathOverflow, November 2011.
- [13] Doug Hensley. Metric Diophantine Approximation and Probability. *New York Journal of Mathematics*, 4: 249-257, July 1998.
- [14] Viviane Baladi and Brigitte Valle. Euclidean Algorithms Are Gaussian. *Journal of Number Theory*, 110(2): 331-386, February 2005.
- [15] Donald E. Knuth. *Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Pearson Education, 2014.

- [16] Wikipedia contributors. Lochs's theorem, 2024. [Online; accessed 14-October-2024].
- [17] Giedrius Alkauskas. Transfer Operator for the Gauss Continued Fraction Map - Structure of the Eigenvalues and Trace Formulas, 2018.
- [18] George J. Klir and Bo Yuan. Fuzzy Sets and Fuzzy Logic - Theory and Applications. Prentice Hall PTR, 1995.
- [19] John James Milne. An Elementary Treatise on Cross-ratio Geometry: With Historical Notes. The University Press, 1911.
- [20] P. M. Williams. Bayesian Conditionalisation and the Principle of Minimum Information. *The British Journal for the Philosophy of Science*, 31(2): 131-144, 1980.
- [21] Wikipedia contributors, the Free Encyclopedia. Poincaré disk model, 2023. [Online; accessed 20-January-2023].
- [22] Linas Vepstas. The Gauss-Kuzmin-Wirsing Operator. Unpublished, 2012. Revised 7 July 2020.
- [23] Julio Rives. The Code Underneath. (Submitted to) *Axioms*, pages 1-60, 2024.