

Article

# Taking Rational Numbers at Random

Nicola Cufaro Petroni <sup>†</sup>

Department of Mathematics and TIRES, University of Bari, INFN Sezione di Bari, via E. Orabona 4, 70125 Bari, Italy; cufaro@ba.infn.it

<sup>†</sup> The author, Nicola Cufaro Petroni, died in the last week of June 2023. The paper had been seen by four reviewers who expressed interest for publication, and asked for some improvements. The author did not have the possibility of writing a new version. Giovanni M. Cicuta agreed to follow up with the publication process. The paper is unchanged (except for editorial reasons) and a new Section 8 was added.

**Abstract:** In this article, some prescriptions to define a distribution on the set  $\mathbb{Q}_0$  of all rational numbers in  $[0, 1]$  are outlined. We explored a few properties of these distributions and the possibility of making these rational numbers *asymptotically equiprobable* in a suitable sense. In particular, it will be shown that in the said limit—albeit no absolutely continuous uniform distribution can be properly defined in  $\mathbb{Q}_0$ —the probability allotted to every single  $q \in \mathbb{Q}_0$  asymptotically vanishes, while that of the subset of  $\mathbb{Q}_0$  falling in an interval  $[a, b] \subseteq \mathbb{Q}_0$  goes to  $b - a$ . We finally present some hints to complete sequencing without repeating the numbers in  $\mathbb{Q}_0$  as a prerequisite to laying down more distributions on it.

**Keywords:** rational numbers; discrete distributions; randomness

## 1. Introduction

What could the locution *taking at random* possibly mean? In its most general sense, this would indicate that the drawing of an outcome  $\omega$  out of a set  $\Omega$  is made according to any arbitrary (but legitimate) probability measure assigned on the subsets of  $\Omega$ , and then that the usual precepts of the probability theory are followed by the result that different probabilities are normally allotted to distinct subsets of  $\Omega$ . Traditionally, however, the meaning of the said locution is more circumscribed and stands rather for assuming that there is no reason to think that there are preferred outcomes  $\omega \in \Omega$ , these instead being supposed to be equally likely. This is the meaning that we are interested in within this paper, or—if this notion is not exactly applicable—an asymptotic version of it in some acceptable limiting sense.

It is well known that for the sets of real numbers, this kind of *randomness* is enforced either by sheer *equiprobability* (on the finite sets) or by distribution *uniformity* (on the bounded, Lebesgue measurable, uncountable sets). On the other hand, infinite, countable sets and unbounded, uncountable sets are both excluded from these egalitarian probability attributions because their elements can be made neither equiprobable (with a non-vanishing probability, in the countable case), nor uniformly distributed (with a non-vanishing probability density, in the uncountable case). On these occasions, it is therefore advisable to start with some proper (neither equiprobable, nor uniform) probability distribution, and then to inquire if and how this can be made ever closer—in a suitable, approximate sense—either to an equiprobable distribution or to an uniform one. We will then, respectively, speak of *asymptotic equiprobability* and *asymptotic uniformity*.

The focus of the present inquiry, as will be elucidated in Section 2, are the rational numbers that—even in a bounded interval—constitute an infinite, countable set, with a few relevant, additional peculiarities due to them also being dense everywhere among the real numbers. In Section 3, we provide a procedure to attribute non-vanishing probabilities to every rational number  $q = n/m$  in the interval  $[0, 1]$ . Section 4 is instead devoted to the



**Citation:** Cufaro Petroni, N. Taking Rational Numbers at Random. *AppliedMath* **2023**, *3*, 648–663. <https://doi.org/10.3390/appliedmath3030034>

Academic Editor: Darin J. Ulness

Received: 18 April 2023

Revised: 1 August 2023

Accepted: 11 August 2023

Published: 1 September 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

aftermath of supposing conditionally equiprobable numerators  $n$ , and then Section 5 shows under which hypotheses our distributions can give rise to an asymptotic equiprobability of the rationals in  $[0, 1]$  such that—without pretending to have an absolutely continuous (ac) uniform distribution on  $\mathbb{Q}_0$ —the probability allotted to a single  $q \in \mathbb{Q}_0$  vanishes in the limit, while that of the infinite subset of rationals falling in an interval  $[a, b] \subseteq \mathbb{Q}_0$  goes to  $b - a$ . Several examples of denominator distributions are explained in Section 6, giving rise to a few closed formulas. Finally, in Section 7, some concluding remarks are added, offering a glimpse into the open problem of sequencing all the rational numbers in  $\mathbb{Q}_0$ .

## 2. Probability on Rational Numbers

Rational numbers  $\mathbb{Q}$  are famously countable, and hence they can be put in a sequence. Since they are a dense subset of the real numbers, every rational number is a cluster point, and thus no sequence encompassing *all of them* can ever converge, not to say be monotone. In any case, their countability certainly allows the allotment of discrete distributions with non-vanishing probabilities to *every* rational number; since they are infinite, however, they can never be exactly *equiprobable*. We will outline in the forthcoming sections a simple procedure to obtain distributions on all the rationals in  $[0, 1]$ , a set that we will shortly denote as  $\mathbb{Q}_0 = \mathbb{Q} \cap [0, 1]$ , and we will investigate if and how they can be considered *asymptotically equiprobable*. We will refrain, for the time being, from extending these considerations to the whole of  $\mathbb{Q}$  only because in our opinion, this would not add particular insight into the discussion at the present stage of the inquiry.

It is, however, advisable to remember right away that the distribution of a *rv* (random variable)  $Q$  taking values in  $\mathbb{Q}_0$  must anyhow be of a discrete type, allotting (possibly non-vanishing) probabilities to the individual rational numbers  $q \in \mathbb{Q}_0$ : conceivable continuous set functions—namely with continuous, albeit perhaps not absolutely continuous, *cdf* (cumulative distribution function)—would turn out not to be countably additive, and would hence not qualify as measures, not to say as probability distributions. Every continuous *cdf* for  $Q$  would indeed entail that, at the same time,  $P\{Q = q\} = 0, \forall q \in \mathbb{Q}_0$ , and  $P\{Q \in \mathbb{Q}_0\} = 1$ , while  $\mathbb{Q}_0$  is apparently the countable union of the disjoint, negligible sets  $\{q\}$ , which is in plain conflict with the countable additivity. This in particular also rules out, for the numbers in  $\mathbb{Q}_0$ , the possibility of being in some sense *uniformly distributed* (an imaginable surrogate of equiprobability suggested by the cited density of the rationals) for the numbers in  $\mathbb{Q}_0$ . This property would, in fact, require a *cdf* of the uniform type for  $Q$ :

$$F_Q(x) = P\{Q \leq x\} = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

which is apparently continuous, and would hence attribute probability 0 to every single  $q$ , but probability 1 to  $\mathbb{Q}_0$ .

We would like to stress, moreover, that the problem focused on in the present paper is not how to realistically produce rational numbers that are possibly equiprobable *at random*; this would be performed trivially, for instance, just by taking random, uniformly distributed *real* numbers and then truncating them to a prefixed number  $n$  of decimal digits, as always conducted in practice in every computer simulation of random numbers in  $[0, 1]$ . It is indeed apparent that in so doing, we would shrink  $\mathbb{Q}_0$  to a *finite* set of rational numbers (they would be exactly  $10^n + 1$ ) that could always be made exactly equiprobable, failing on the other hand to allot a non-vanishing probability to the remaining, overwhelmingly more numerous elements of  $\mathbb{Q}_0$ . The aim of our inquiry is instead to find a sensible way to attribute (non-vanishing and possibly not too different from each other) probabilities *to every rational number* in  $\mathbb{Q}_0$ . Their practical simulation is not considered our main purpose here, but just as an eventual side effect of this allocation.

Remark that one could be lured to think that a way around the previous snag could consist of again drawing again uniformly distributed real numbers, yet truncating the decimal digits to some *random* number  $N$  taking arbitrary, finite but unbounded integer

values. Even in this way, however, not every rational number would have a chance to be produced; the said procedure would indeed a priori exclude all the (infinitely many) rational numbers with an infinite, periodic decimal representation, for instance,  $1/3, 2/3, \dots$  and so on. In light of this preliminary scrutiny, the best way to tackle the task of laying down a probability on  $\mathbb{Q}_0$  seems then to be to exploit the fractional representation  $q = n/m$  of every rational number by attributing some suitable joint distribution to its numerators and denominators, considered here as *rvs* with integer values.

### 3. Distributions on $\mathbb{Q}_0$

Using the well-known diagram used to show how rational numbers are countable, we will explore two dependent *rvs*,  $M$  and  $N$ , with integer values

$$m = 1, 2, \dots \qquad n = 0, 1, 2, \dots, m$$

acting, respectively, as denominator and numerator of the random rational number  $Q = N/M \in [0, 1]$ . Therefore,  $Q$  will have the values  $q = n/m$  arrayed in a triangular scheme as in Table 1.

**Table 1.** Table of rational numbers  $q = n/m$  with repetitions: many fractions are reducible to canonical forms already present in earlier positions.

		<i>n</i>									
<i>m</i>	0	1	2	3	4	5	6	7	8	...	
1	0	1									
2	0	1/2	1								
3	0	1/3	2/3	1							
4	0	1/4	2/4	3/4	1						
5	0	1/5	2/5	3/5	4/5	1					
6	0	1/6	2/6	3/6	4/6	5/6	1				
7	0	1/7	2/7	3/7	4/7	5/7	6/7	1			
8	0	1/8	2/8	3/8	4/8	5/8	6/8	7/8	1		
⋮	⋮									⋱	

However, this method causes every rational number  $q$  to appear infinitely many times because of the existence of reducible fractions. For example, using the usual notation for *repeating decimals*, we have

$$0.5 = 1/2 = 2/4 = 3/6 = \dots \qquad 0.\bar{3} = 1/3 = 2/6 = \dots \qquad 0.75 = 3/4 = 6/8 = \dots$$

to avoid repetitions, the rational numbers in  $[0, 1]$  should be listed with blanks as in Table 2. However, this table is not suitable for assigning probabilities directly to its elements, as there is no simple way to give them a sequential index (for example, which one is the 1000th element?). This is because the numbers  $v_m$  of different rationals in each row with the same irreducible denominator  $m$  form a rather irregular sequence, as we will briefly explain in Section 7. Therefore, it is better to use the complete Table 1 and introduce a joint distribution of  $N$  and  $M$  first.

$$\begin{aligned} P\{M = m\} & \qquad m = 1, 2, \dots \\ P\{N = n | M = m\} & \qquad n = 0, 1, 2, \dots, m \\ P\{N = n, M = m\} & = P\{N = n | M = m\} P\{M = m\} \end{aligned}$$

**Table 2.** Table of rational numbers  $q \doteq n/m$  without repetitions: only irreducible fractions are represented, along with the number  $v_m$  of the different rationals sharing a common irreducible denominator  $m$ .

		$n$								
$v_m$	$m$	0	1	2	3	4	5	6	7	...
2	1	0	1							
1	2		$1/2$							
2	3		$1/3$	$2/3$						
2	4		$1/4$		$3/4$					
4	5		$1/5$	$2/5$	$3/5$	$4/5$				
2	6		$1/6$				$5/6$			
6	7		$1/7$	$2/7$	$3/7$	$4/7$	$5/7$	$6/7$		
4	8		$1/8$		$3/8$		$5/8$		$7/8$	
$\vdots$	$\vdots$		$\vdots$							$\ddots$

For a rational number  $q$ , we adopt the notation

$$q \doteq n/m$$

to indicate that  $n/m$  is the irreducible representation of  $q$ , namely that  $n$  and  $m$  are co-primes. For instance, in the previous examples:

$$0.5 \doteq 1/2 \quad 0.\bar{3} \doteq 1/3 \quad 0.75 \doteq 3/4$$

To account for the repeated entries in Table 1, we will allot the probability (the notation used in this paper follows the usual one of probability textbooks.) to every rational  $q \doteq n/m \in [0, 1]$ .

$$\begin{aligned} P\{Q = q\} &= \sum_{\ell=1}^{\infty} P\{N = \ell n, M = \ell m\} \\ &= \sum_{\ell=1}^{\infty} P\{N = \ell n | M = \ell m\} P\{M = \ell m\} \end{aligned} \tag{1}$$

This apparently also defines the *cdf* of  $Q$  as (here, of course,  $x \in \mathbb{R}$ )

$$\begin{aligned} F_Q(x) &= P\{Q \leq x\} = P\{N \leq Mx\} = \sum_{m=1}^{\infty} P\{N \leq mx | M = m\} P\{M = m\} \\ &= \sum_{m=1}^{\infty} F_N(mx | M = m) P\{M = m\} \end{aligned} \tag{2}$$

and the probability of  $Q$  falling in  $(a, b]$ , for  $0 \leq a < b \leq 1$  real numbers, as:

$$\begin{aligned} P\{a < Q \leq b\} &= F_Q(b) - F_Q(a) \\ &= \sum_{m=1}^{\infty} [F_N(mb | M = m) - F_N(ma | M = m)] P\{M = m\} \end{aligned} \tag{3}$$

Notice that the conditional *cdf* of  $N$  can also be given as

$$\begin{aligned} F_N(x | M = m) &= P\{N \leq x | M = m\} = \sum_{n=0}^m P\{N = n | M = m\} \theta(x - n) \\ &= \sum_{n=0}^{\lfloor x \rfloor} P\{N = n | M = m\} \end{aligned} \tag{4}$$

where

$$\vartheta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

is the Heaviside function, while for every real number  $x$ , the symbol  $\lfloor x \rfloor$  denotes the floor of  $x$ , namely the greatest integer less than or equal to  $x$ . Therefore, Equations (2) and (3) can be written in the following form:

$$F_Q(x) = \sum_{m=1}^{\infty} P\{M = m\} \sum_{n=0}^{\lfloor mx \rfloor} P\{N = n | M = m\} \tag{5}$$

$$P\{a < Q \leq b\} = \sum_{m=1}^{\infty} P\{M = m\} (1 - \delta_{\lfloor ma \rfloor, \lfloor mb \rfloor}) \sum_{n=\lfloor ma \rfloor + 1}^{\lfloor mb \rfloor} P\{N = n | M = m\} \tag{6}$$

where the Kronecker delta considers the circumstance that when  $\lfloor mb \rfloor = \lfloor ma \rfloor$ , the expression becomes vanishing, resulting in  $\lfloor mb \rfloor \geq \lfloor ma \rfloor + 1$ . Additionally, we possess expressions for both the expectations and the characteristic function:

$$E[Q] = E[N/M] = E\left[\frac{1}{M} E[N|M]\right] = \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m} E[N|M = m] \tag{7}$$

$$E[Q^2] = \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m^2} E[N^2|M = m] \tag{8}$$

$$\begin{aligned} \varphi_Q(u) &= E\left[e^{iuN/M}\right] = \sum_{m=1}^{\infty} P\{M = m\} \sum_{n=0}^m e^{iun/m} P\{N = n | M = m\} \\ &= \sum_{m=1}^{\infty} P\{M = m\} \varphi_N(u/m | M = m) \end{aligned} \tag{9}$$

where we have also introduced the shorthand notation

$$\varphi_N(u | M = m) = E\left[e^{iuN} | M = m\right] = \sum_{n=0}^m e^{iun} P\{N = n | M = m\}$$

The specific joint distributions of variables  $N$  and  $M$  can be chosen in several ways and we survey a few particular cases in the next sections.

#### 4. Equiprobable Numerators

Firstly, we suppose that for a given denominator  $m \geq 1$ , the  $m + 1$  possible values of the numerator  $n = 0, 1, \dots, m$  are equiprobable in the sense that

$$P\{N = n | M = m\} = \frac{1}{m + 1} \quad n = 0, 1, \dots, m$$

We then have (see [1] 0.121)

$$\begin{aligned} E[N | M = m] &= \sum_{n=0}^m \frac{n}{m + 1} = \frac{1}{m + 1} \frac{m(m + 1)}{2} = \frac{m}{2} \\ E[N^2 | M = m] &= \sum_{n=0}^m \frac{n^2}{m + 1} = \frac{1}{m + 1} \frac{m(m + 1)(2m + 1)}{6} = \frac{m(2m + 1)}{6} \end{aligned}$$

and hence, from (7) and (8),

$$\begin{aligned}
 E[Q] &= \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m} \frac{m}{2} = \frac{1}{2} \sum_{m=1}^{\infty} P\{M = m\} = \frac{1}{2} \\
 E[Q^2] &= \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m^2} E[N^2|M = m] = \sum_{m=1}^{\infty} \frac{2m+1}{6m} P\{M = m\} \\
 &= \frac{1}{3} \sum_{m=1}^{\infty} P\{M = m\} + \frac{1}{6} \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m} = \frac{1}{3} + \frac{1}{6} E\left[\frac{1}{M}\right] \\
 V[Q] &= E[Q^2] - E[Q]^2 = \frac{1}{12} + \frac{1}{6} E\left[\frac{1}{M}\right]
 \end{aligned}$$

As for the distribution, with  $n, m$  co-primes and  $0 \leq n \leq m$ , from (1), we have

$$P\{Q = q\} = \sum_{\ell=1}^{\infty} \frac{P\{M = \ell m\}}{\ell m + 1} \quad q \doteq n/m \tag{10}$$

which is independent of  $n$  and relies solely on the value of the irreducible denominator  $m$ . Furthermore, the characteristic Function (9) reads

$$\varphi_N(u|M = m) = \frac{1}{m+1} \sum_{n=0}^m e^{iun} \quad \varphi_Q(u) = \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m+1} \sum_{n=0}^m e^{iun/m}$$

while for the cdf (5), we have, from (4),

$$\begin{aligned}
 F_N(mx|M = m) &= \frac{1}{m+1} \sum_{n=0}^m \vartheta(mx - n) = \begin{cases} 0 & x < 0 \\ \frac{\lfloor mx \rfloor + 1}{m+1} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases} \\
 F_Q(x) &= \sum_{m=1}^{\infty} \frac{P\{M = m\}}{m+1} \sum_{n=0}^m \vartheta(mx - n) \\
 &= \begin{cases} 0 & x < 0 \\ \sum_{m \geq 1} P\{M = m\} \frac{\lfloor mx \rfloor + 1}{m+1} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases} \tag{11}
 \end{aligned}$$

and the probability (6) with  $0 \leq a < b \leq 1$  becomes

$$P\{a < Q \leq b\} = \sum_{m=1}^{\infty} P\{M = m\} \frac{\lfloor mb \rfloor - \lfloor ma \rfloor}{m+1} \tag{12}$$

It is clear that, with the exception of the expected value  $E[Q] = 1/2$ , all these quantities are influenced by the selection of the denominator distribution. However, we will demonstrate in the subsequent section that, given reasonable conditions on the denominators  $M$ , the distribution of  $Q$  can indeed be made arbitrarily close to, though not precisely coincident with, a uniform distribution on the interval  $[0, 1]$ . This behavior is what we will refer to as *asymptotic equiprobability*.

### 5. Asymptotic Equiprobability

Considering the equiprobable numerators introduced in the preceding section, where we denote the distribution of  $M$  as  $p_m = P\{M = m\}$  and the supremum of all its values as  $s$ , let us now consider a sequence of denominators  $M_{k,k \geq 1}$  with corresponding distributions  $p_m(k)_{k \geq 1}$ . Furthermore, let  $s_k$  approach zero as  $k$  tends to infinity in a manner such that

$$\lim_k s_k \ln k = 0 \tag{13}$$

To put it differently, we examine a sequence of distributions that progressively become flatter and approach zero uniformly, consequently leading to increasingly equiprobable denominators. Illustrative examples of such sequences for various values of  $k$  (where  $k$  takes on the values  $1, 2, \dots$ ) include the sequence of *finite equiprobable* distributions

$$p_m(k) = \begin{cases} 1/k & m = 1, 2, \dots, k \\ 0 & m > k \end{cases}$$

where  $s_k = 1/k \xrightarrow{k} 0$  and (13) is satisfied; that of the *geometric* distributions

$$p_m(k) = w_k(1 - w_k)^{m-1} \quad m = 1, 2, \dots$$

with infinitesimal  $w_k$  so that  $s_k = w_k \xrightarrow{k} 0$ : (13) is satisfied with a suitable choice of  $w_k$ ; and finally, that of the *Poisson* distributions

$$p_m(k) = e^{-\lambda_k} \frac{\lambda_k^{m-1}}{(m-1)!} \quad m = 1, 2, \dots$$

with divergent  $\lambda_k$ , where again the modal values are infinitesimal: we know indeed that a Poisson distribution attains its maximum in  $\lfloor \lambda_k \rfloor + 1$ , so that for  $\lambda_k \xrightarrow{k} +\infty$ , its modal value  $s_k$  essentially behaves as (see [1] 8.327.1)

$$s_k = e^{-\lambda_k} \frac{\lambda_k^{\lambda_k-1}}{\Gamma(\lambda_k)} = \frac{1}{\sqrt{2\pi\lambda_k}(1 + O(\lambda_k^{-1}))} \xrightarrow{k} 0$$

In this case, (13) is also fulfilled through a suitable choice of  $\lambda_k$

**Lemma 1.** *Within the aforementioned notations and under the given conditions, we have*

$$\mu_k = E\left[1/M_k\right] = \sum_{m=1}^{\infty} \frac{p_m(k)}{m} \xrightarrow{k} 0 \tag{14}$$

**Proof.** The positive series defining  $\mu_k$  is certainly convergent because

$$\mu_k = \sum_{m=1}^{\infty} \frac{p_m(k)}{m} < \sum_{m=1}^{\infty} p_m(k) = 1$$

and, hence, we can write

$$\mu_k = \sum_{m=1}^{\infty} \frac{p_m(k)}{m} = \sum_{m=1}^k \frac{p_m(k)}{m} + R_k$$

where

$$R_k = \sum_{m=k+1}^{\infty} \frac{p_m(k)}{m} \xrightarrow{k} 0$$

is an infinitesimal remainder. Remark that here  $k$  plays the roles of index of the distribution sequence as well as the cut-off of the series.

On the other hand, under our stated conditions,

$$\sum_{m=1}^k \frac{p_m(k)}{m} < s_k \sum_{m=1}^k \frac{1}{m} = s_k H_k$$

where  $H_k$  denotes the  $k^{th}$  harmonic number, namely the sum of the reciprocal integers up to  $1/k$ . It is well known ([1] 0.131) that, for  $k \rightarrow \infty$ ,  $H_k$  grows as  $\ln k$ , so using (13), we have  $s_k H_k \xrightarrow{k} 0$ , and finally  $\mu_k = s_k H_k + R_k \xrightarrow{k} 0$ .  $\square$

**Theorem 1.** If  $Q = N/M$  and  $F_Q(x)$  is its cdf, then, given the notation and conditions outlined above, we have

$$\lim_k P\{Q = q\} = 0 \qquad \lim_k P\{a < Q \leq b\} = b - a \tag{15}$$

$$\lim_k F_Q(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases} \tag{16}$$

**Proof.** Since our series have positive terms, the first result in (15) follows from (10) and (14) because, with  $q \doteq n/j$ ,

$$P\{Q = q\} = \sum_{\ell=1}^{\infty} \frac{p_{\ell j}(k)}{\ell j + 1} < \sum_{m=1}^{\infty} \frac{p_m(k)}{m + 1} < \sum_{m=1}^{\infty} \frac{p_m(k)}{m} = \mu_k \xrightarrow{k} 0$$

As for the second result in (15), since for every real number  $x$  it is  $x - 1 \leq \lfloor x \rfloor \leq x$ , for every  $k = 1, 2, \dots$ , and  $0 \leq a < b \leq 1$ , we have, from (12),

$$\sum_{m=1}^{\infty} p_m(k) \frac{m(b-a) - 1}{m + 1} \leq P\{a < Q \leq b\} \leq \sum_{m=1}^{\infty} p_m(k) \frac{m(b-a) + 1}{m + 1}$$

namely,

$$b - a + (a - b - 1) \sum_{m=1}^{\infty} \frac{p_m(k)}{m + 1} \leq P\{a < Q \leq b\} \leq b - a + (a - b + 1) \sum_{m=1}^{\infty} \frac{p_m(k)}{m + 1}$$

so that, since  $a - b - 1 \leq 0$  and  $a - b + 1 \geq 0$ , it is

$$b - a + (a - b - 1)\mu_k \leq P\{a < Q \leq b\} \leq b - a + (a - b + 1)\mu_k$$

and the second result (15) follows again from (14). Finally, in a similar way, we find, for (16), that

$$\sum_{m=1}^{\infty} p_m(k) \frac{mx}{m + 1} \leq F_Q(x) \leq \sum_{m=1}^{\infty} p_m(k) \frac{mx + 1}{m + 1} \qquad 0 \leq x \leq 1$$

namely,

$$x - x \sum_{m=1}^{\infty} \frac{p_m(k)}{m + 1} \leq F_Q(x) \leq x + (1 - x) \sum_{m=1}^{\infty} \frac{p_m(k)}{m + 1}$$

and, hence,

$$x - x \mu_k < F_Q(x) < x + (1 - x)\mu_k$$

so that, in this case, the result also follows from (14).  $\square$

This theorem reveals that as the upper limit  $k$  approaches infinity; although the probability associated with individual rational numbers tends to vanish, the probability of these numbers grouped into intervals remains substantial. This behavior bears a striking resemblance to the behavior observed in continuously distributed *real rvs*. Nevertheless, due to the factors discussed in Section 2, it is important to emphasize that the previous result does not imply the possibility of achieving a uniform limit distribution on  $\mathbb{Q}_0$  (as such a distribution does not exist). Instead, it suggests that our stochastic rational numbers  $Q$ —at least for denominators  $m$  distributed in a fairly flat way, and numerators  $n$  that are



conditionally equiprobable within the range of 0 to  $m$ —tend to behave asymptotically like a uniform distribution within the interval  $[0, 1]$ . This aligns well with our intuitive concept of *selecting rational numbers randomly*. It is worth noting in this context that the variance is also given by

$$V[Q] = \frac{1}{12} + \frac{1}{6} E\left[\frac{1}{M}\right] = \frac{1}{12} + \frac{\mu_k}{6} \xrightarrow{k} \frac{1}{12}$$

again in agreement with an approximate uniform distribution in  $[0, 1]$ .

### 6. Denominator Distributions

#### 6.1. Geometric Denominators

A few closed formulas about the *rv*  $Q$  are available for particular denominator distributions: let us suppose, for instance, that the denominator  $M$  is geometrically distributed as

$$P\{M = m\} = w(1 - w)^{m-1} \quad w > 0, \quad m = 1, 2, \dots$$

In this case, we first find

$$E[M] = \frac{1}{w} \quad E\left[\frac{1}{M}\right] = \sum_{m=1}^{\infty} \frac{w(1 - w)^{m-1}}{m} = \frac{w}{1 - w} \sum_{m=1}^{\infty} \frac{(1 - w)^m}{m} = -\frac{w \ln w}{1 - w}$$

and, hence,

$$V[Q] = \frac{1}{12} - \frac{w \ln w}{6(1 - w)}$$

while for the *cdf* we can not go beyond its formal definition

$$F_Q(x) = \begin{cases} 0 & x < 0 \\ \sum_{m \geq 1} w(1 - w)^{m-1} \frac{\lfloor mx \rfloor + 1}{m+1} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

As for the  $Q$  discrete distribution instead, taking  $q \doteq n/m$ , we find with  $j = \ell - 1$  the analytic expression

$$\begin{aligned} P\{Q = q\} &= \sum_{\ell=1}^{\infty} \frac{w(1 - w)^{\ell m - 1}}{\ell m + 1} = w(1 - w)^{m-1} \sum_{\ell=1}^{\infty} \frac{(1 - w)^{m(\ell-1)}}{\ell m + 1} \\ &= w(1 - w)^{m-1} \sum_{j=0}^{\infty} \frac{(1 - w)^{mj}}{m(j+1) + 1} \\ &= \frac{w(1 - w)^{m-1}}{m + 1} {}_2F_1\left(1, 1 + \frac{1}{m}; 2 + \frac{1}{m}; (1 - w)^m\right) \end{aligned} \tag{17}$$

where  ${}_2F_1(a, b; c; z)$  is a hypergeometric function [1] that quantifies the deviation of  $P\{Q = q\}$  from the corresponding joint probability of  $N, M$

$$P\{N = n, M = m\} = \frac{w(1 - w)^{m-1}}{m + 1}$$

This formula allows a graphic representation of  $P\{Q = q\}$  as a function of the irreducible denominators  $m$  displayed in Figure 1, where it is clear how the initial ( $m = 1$ ) ordering of the probabilities (increasing with the  $w$  values going from  $w = 0.001$  to  $w = 0.9$ ) becomes totally overturned for a large enough  $m$ . We remark that each value of the probability (17) should be considered to be attributed to every rational number with the same  $m$  as the irreducible denominator. For instance, see Table 2; for  $m = 1$ , we obtain the probability of  $q \doteq 0$  and 1; for  $m = 2$ , the probability of  $q \doteq 1/2$  alone; for  $m = 3$ , that of  $q \doteq 1/3, 2/3$ ; for  $m = 4$ , that of  $q \doteq 1/4, 3/4$ ; ...; and so on. This observation enables us,

in particular, to avoid a potential misinterpretation. Indeed, we note that, while it might seem at first glance that

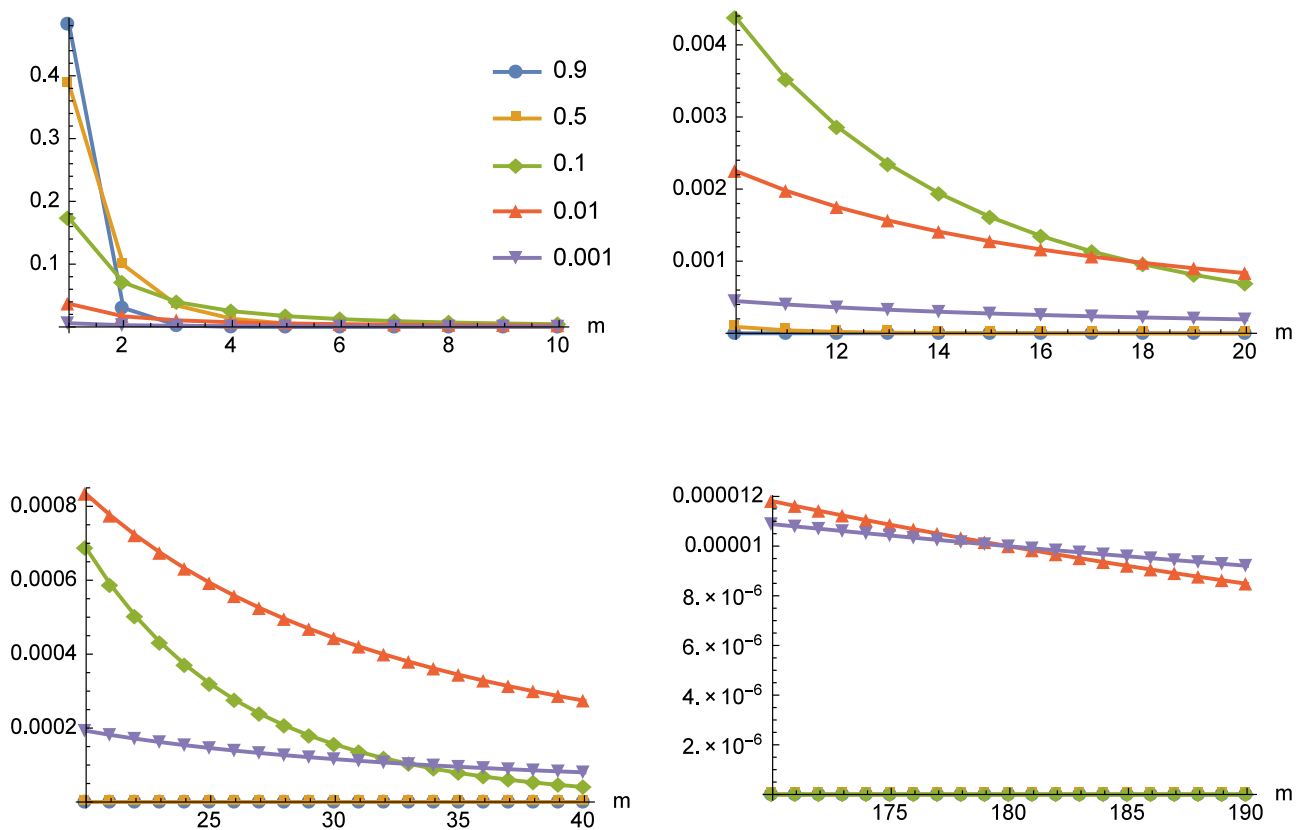
$$\sum_{m=1}^{\infty} \sum_{n=0}^m P\{N = n, M = m\} = \sum_{m=1}^{\infty} \sum_{n=0}^m \frac{w(1-w)^{m-1}}{m+1} = \sum_{m=1}^{\infty} w(1-w)^{m-1} = 1$$

we find instead

$$\sum_{m=1}^{\infty} \sum_{n=0}^m \frac{w(1-w)^{m-1}}{m+1} {}_2F_1\left(1, 1 + \frac{1}{m}; 2 + \frac{1}{m}; (1-w)^m\right) < 1$$

as can be seen from the fact that for  $0 < w < 1$

$${}_2F_1\left(1, 1 + \frac{1}{m}; 2 + \frac{1}{m}; (1-w)^m\right) \begin{cases} = 1 & m = 1 \\ < 1 & m = 2, 3, \dots \end{cases}$$



**Figure 1.** Probabilities (17) attributed to rational numbers as a function of the irreducible, geometrically distributed denominators  $m$ , and for decreasing (0.9, 0.5, 0.1, 0.01, 0.001) values of  $w$ : by choosing different  $m$  intervals, the pictures show how these probabilities level down to infinitesimal equiprobability for  $w \rightarrow 0$ .

This is not in contradiction with the mandatory requirement that

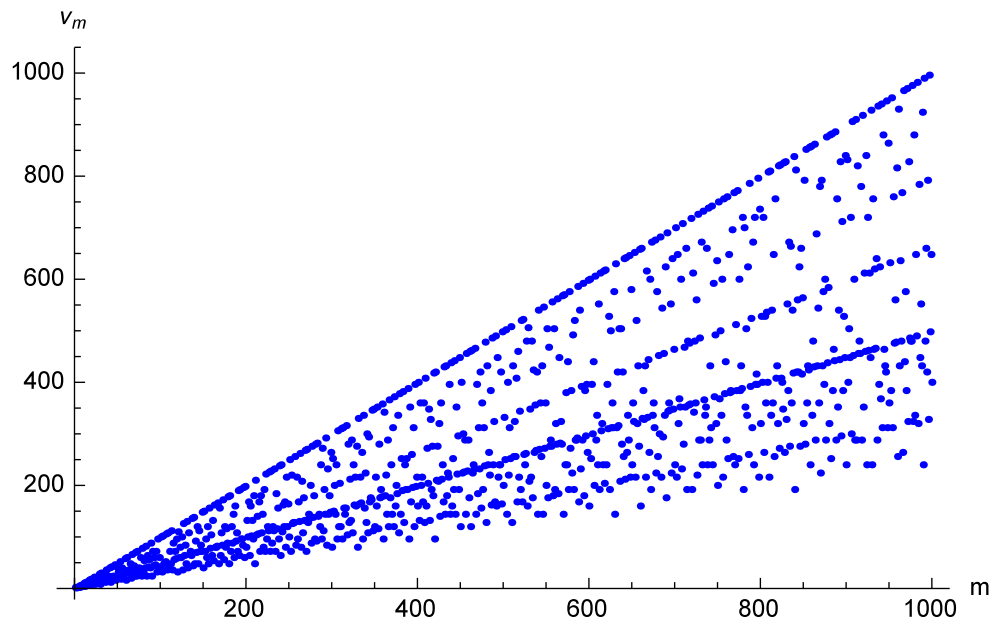
$$\sum_{q \in \mathbb{Q}_0} P\{Q = q\} = 1 \tag{18}$$

precisely because, as previously remarked, the probability associated to an  $m$  must be attributed to several different rational numbers  $q$ ; if  $v_m$  is the number of rationals  $q$  that

have  $m$  as its irreducible denominator, then we should rather pay attention to ascertain the normalization in the form

$$\begin{aligned} \sum_{q \in \mathbb{Q}_0} P\{Q = q\} &= \sum_{m=1}^{\infty} \sum_{n=0}^m v_m \frac{w(1-w)^{m-1}}{m+1} {}_2F_1\left(1, 1 + \frac{1}{m}; 2 + \frac{1}{m}; (1-w)^m\right) \\ &= \sum_{m=1}^{\infty} v_m w(1-w)^{m-1} {}_2F_1\left(1, 1 + \frac{1}{m}; 2 + \frac{1}{m}; (1-w)^m\right) = 1 \end{aligned}$$

We stress that confirming this result—which we can confidently establish through construction and definition—proves to be challenging when attempting direct calculations. This difficulty arises due to the lack of a readily available closed-form expression for the sequence  $v_m$ . The behavior of this sequence is rather irregular, though it demonstrates an overall upward trend on average. This can be observed from an empirical plot of its initial values, depicted in Figure 2. We will defer a more detailed discussion on this matter to Section 7, where we will provide additional insights. In particular, we will demonstrate how the preceding normalization condition can be employed for stepwise computations of the  $v_m$  values.



**Figure 2.** Numerosity  $v_m$  of the different rational numbers  $q \doteq n/m$  sharing a common, irreducible denominator  $m$ .

### 6.2. Poisson and Equiprobable Denominators

When the denominators are distributed according to different (albeit simple) laws, we no longer have access to elementary closed forms for  $P\{Q = q\}$ . For instance, if  $M$  follows a Poisson distribution, we have

$$P\{M = m\} = e^{-\lambda} \frac{\lambda^{m-1}}{(m-1)!} \quad \lambda > 0, \quad m = 1, 2, \dots$$

we find  $E[M] = 1 + \lambda$  and

$$E\left[\frac{1}{M}\right] = \sum_{m=1}^{\infty} \frac{e^{-\lambda}}{m} \frac{\lambda^{m-1}}{(m-1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{m!} = \frac{e^{-\lambda}}{\lambda} \left( \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} - 1 \right) = \frac{1 - e^{-\lambda}}{\lambda}$$

while for the variance, we have

$$V[Q] = \frac{1}{12} + \frac{1 - e^{-\lambda}}{6\lambda}$$

but the *cdf* is

$$F_Q(x) = \begin{cases} 0 & x < 0 \\ e^{-\lambda} \sum_{m \geq 1} \frac{\lambda^{m-1}}{(m-1)!} \frac{\lfloor mx \rfloor + 1}{m+1} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

and for the discrete distribution, taking  $q \doteq n/m$ , we find

$$P\{Q = q\} = \sum_{\ell=1}^{\infty} \frac{e^{-\lambda}}{\ell m + 1} \frac{\lambda^{\ell m - 1}}{(\ell m - 1)!}$$

with no closed expression readily available.

Considering, instead, denominators  $M$  taking only a finite number  $k = 1, 2, \dots$  of equiprobable values  $m$ , we have:

$$P\{M = m\} = \begin{cases} 1/k & m = 1, 2, \dots, k \\ 0 & m > k \end{cases}$$

We then have

$$E[M] = \frac{k+1}{2} \quad E\left[\frac{1}{M}\right] = \frac{1}{k} \sum_{m=1}^k \frac{1}{m} = \frac{H_k}{k}$$

and, hence,

$$V[Q] = \frac{1}{12} + \frac{H_k}{6k}$$

while for the *cdf* it is

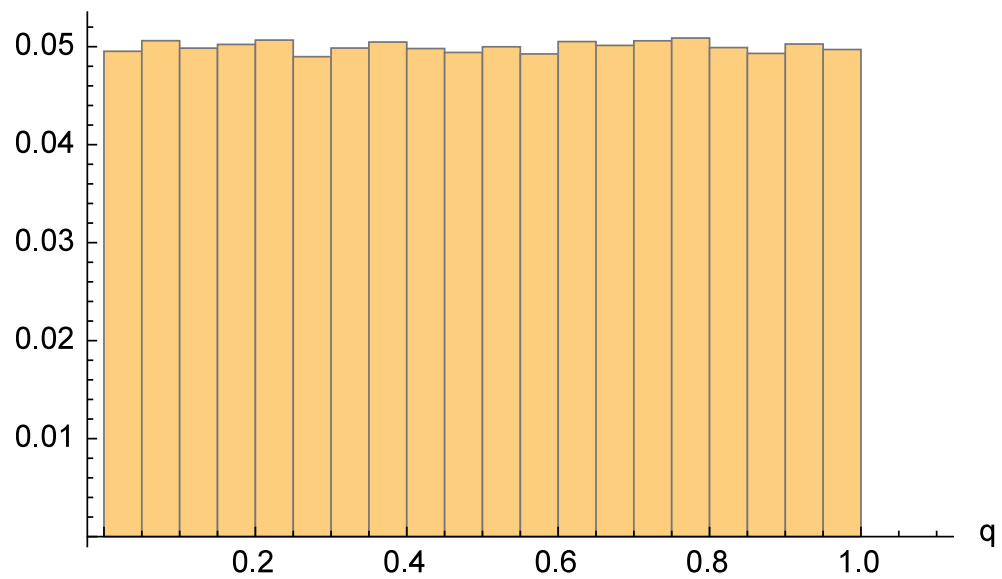
$$F_Q(x) = \frac{1}{k} \sum_{m=1}^k \frac{\lfloor mx \rfloor + 1}{m+1}$$

and the discrete distribution probabilities are

$$P\{Q = q\} = \frac{1}{k} \sum_{\ell=1}^{\lfloor k/m \rfloor} \frac{1}{\ell m + 1} \quad q \doteq n/m$$

where, since  $m \leq k$ , it is always  $\lfloor k/m \rfloor = 1, 2, \dots$

Even in this case, we have no closed formulas to show. Since the sums involved are now always finite, a simple—but essentially trivial—technique for simulating an *asymptotically equiprobable sample* of rational numbers within the interval  $[0, 1]$  seems to emerge. The approach involves several steps. Begin by selecting a sufficiently large value  $k$ . Then, randomly select an integer  $m$  from the equiprobable set of numbers  $1, 2, \dots, k$ . Subsequently, choose another random integer  $n$  from the equiprobable set of numbers  $0, 1, \dots, m$ , and calculate  $q = n/m$ . By repeating this process a substantial number of times, an almost uniformly distributed sample within the range  $[0, 1]$  can be generated, as depicted in Figure 3. Nevertheless, it is important to note that the approach has limitations, as mentioned earlier in Section 2. Not every rational number in  $\mathbb{Q}_0$  would have an opportunity to be drawn using this method, as only a finite subset of these numbers would be taken into account. Although, in theory, this finite set of numbers could be made exactly equiprobable, the infinitely numerous remaining rational numbers would be entirely excluded and assigned a probability of precisely zero.



**Figure 3.** Typical histogram of the relative frequencies of a sample of  $10^5$  random rationals generated following the procedure described in Section 6.2: here, the maximum value of the equiprobable denominators is chosen to be  $k = 10^5$ .

**7. Final Remarks: Sequencing Rational Numbers**

Other examples of distributions on the rational numbers in  $[0, 1]$  are, of course, possible. For instance, given  $0 < p < 1$  and for a given denominator  $m = 1, 2, \dots$ , it is possible to suppose that the numerators are binomially—instead of equiprobably—distributed as

$$P\{N = n | M = m\} = \binom{m}{n} p^n (1 - p)^{m-n} \quad n = 0, 1, \dots, m$$

By choosing then a suitable distribution for the denominator  $M$ , we can define the global distribution of  $Q = N/M$ . However, rather than indulging in displaying these further examples, we would like to conclude this paper with a few remarks about a particular residual open problem.

Due to the countable nature of  $\mathbb{Q}_0$ , as we have already said, its elements within the range  $0 \leq q \leq 1$  can be systematically organized into a sequence. To simplify the process of assigning a distribution to  $\mathbb{Q}_0$ , it would be immensely advantageous if this sequence could encompass all rational numbers within  $[0, 1]$  without any repetitions. Achieving this arrangement, however, requires identifying patterns that govern such sequence. This, in turn, would enable us to easily determine both the rational number  $q$  associated with a given index  $k$  and, conversely, the index  $k$  corresponding to a specific rational number  $q$ . Yet, this endeavor is complicated by the inherent irregularity present in the sequence of entries in the triangular table, such as the unpredictable occurrences of prime numbers among the denominators  $m$ . Notably, even the presence of prime numbers—which uniquely identify rows without gaps between the extremes—does not follow a readily discernible pattern. Nonetheless, it is evident that effectively sequencing all numbers in  $\mathbb{Q}_0$  hinges on obtaining insights into  $v_m$ —specifically, the count of non-blank entries in the  $m$ th row of Table 3.

**Table 3.** Table of rational numbers  $q \doteq n/m$  without repetitions, along with the progressive number  $\nu_m$  of the different rationals sharing a common irreducible denominator  $m$ , and their sums  $\sigma_m$ .

$\sigma_m$	$\nu_m$	$m$	$n$												
			0	1	2	3	4	5	6	7	8	9	10		
1	2	1	0	1											
$1/2$	1	2		$1/2$											
1	2	3		$1/3$	$2/3$										
1	2	4		$1/4$		$3/4$									
2	4	5		$1/5$	$2/5$	$3/5$	$4/5$								
1	2	6		$1/6$				$5/6$							
3	6	7		$1/7$	$2/7$	$3/7$	$4/7$	$5/7$	$6/7$						
2	4	8		$1/8$		$3/8$		$5/8$		$7/8$					
3	6	9		$1/9$	$2/9$		$4/9$	$5/9$		$7/9$	$8/9$				
2	4	10		$1/10$		$3/10$				$7/10$		$9/10$			
5	10	11		$1/11$	$2/11$	$3/11$	$4/11$	$5/11$	$6/11$	$7/11$	$8/11$	$9/11$	$10/11$		
$\vdots$	$\vdots$	$\vdots$		$\vdots$											$\vdots$

While not attempting an exhaustive treatment of this topic, we will focus on several observations concerning some basic properties of the numbers  $\nu_m$  (representing the count of rational numbers with a common irreducible denominator  $m$  in a row of Table 3) and  $\sigma_m$  (indicating the sum of these rational numbers). To begin, it is worth noting that the normalization condition (18) can be employed to derive a systematic method for iteratively computing the values of  $\nu_m$ . For example, as detailed in Section 6.1, when denominators follow a geometric distribution and numerators are conditionally equiprobable, the distribution of  $Q$  is expressed by (17). To satisfy the normalization (18), the number  $\nu_m$  of equiprobable numbers sharing the same irreducible denominator must be taken into consideration. It is apparent that by setting  $z = 1 - w$  in (17), the normalization condition (18) transforms into

$$\sum_{q \in \mathbb{Q}_0} P\{Q = q\} = \frac{1 - z}{z^2} \sum_{m=1}^{\infty} \nu_m \sum_{\ell=1}^{\infty} \frac{z^{m\ell+1}}{m\ell + 1} = 1$$

namely, with a power expansion

$$\sum_{m=1}^{\infty} \nu_m \sum_{\ell=1}^{\infty} \frac{z^{m\ell+1}}{m\ell + 1} = \frac{z^2}{1 - z} = \sum_{j=0}^{\infty} z^{j+2} \tag{19}$$

This relation can be used to find the values of  $\nu_m$  by equating the coefficients of the identical powers of  $z$ . Indeed, writing the first terms of (19), we find

$$\begin{aligned} & \nu_1 \left( \frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \frac{z^5}{5} + \dots \right) + \nu_2 \left( \frac{z^3}{3} + \frac{z^5}{5} + \frac{z^7}{7} + \frac{z^9}{9} + \dots \right) \\ & + \nu_3 \left( \frac{z^4}{4} + \frac{z^7}{7} + \frac{z^{10}}{10} + \frac{z^{13}}{13} + \dots \right) + \nu_4 \left( \frac{z^5}{5} + \frac{z^9}{9} + \frac{z^{13}}{13} + \frac{z^{17}}{17} + \dots \right) + \dots \\ & = z^2 + z^3 + z^4 + z^5 + \dots \end{aligned}$$

and, hence, we progressively have

$$\begin{array}{ll}
 v_1/2 = 1 & v_1 = 2 \\
 v_1/3 + v_2/3 = 1 & v_2 = 1 \\
 v_1/4 + v_3/4 = 1 & v_3 = 2 \\
 v_1/5 + v_2/5 + v_4/5 = 1 & v_4 = 2 \\
 \dots & \dots
 \end{array}$$

and so on, in agreement with the corresponding entries of Table 3. It is crucial to emphasize that this procedure cannot rely on the specific distribution of  $Q$  since the sequence  $v_m$  remains constant, and the normalization condition (18) must hold for any valid distribution.

To conclude, we will outline a few elementary properties of  $v_m$  and  $\sigma_m$  that can be instrumental for future advancements. Here,  $m = 1, 2, \dots$  represents the denominators and  $n = 0, 1, \dots, m$  represents the numerators. We term them *accepted* when  $n/m$  appears in Table 3, meaning it is an irreducible fraction. Here are the properties:

1.  **$v_m \leq m - 1$  for  $m \geq 2$ :** in our table  $n = 0$  and  $n = m$  are accepted only for  $m = 1$  so that, in every row with  $m \geq 2$ , the first and last number are always missing; then, apparently,  $v_m = (m + 1) - 2 = m - 1$ ; in particular,  **$v_m = m - 1$  only, for  $m$  prime number.**
2. **For  $m \geq 3$ , if  $n = k \geq 1$  is accepted, then  $n = m - k \leq m - 1$  is also accepted** because, if  $k/m$  is irreducible, then  $(m-k)/m = 1 - k/m$  is also irreducible, namely, the accepted values always show up in *pairs*; in particular, since  $n = 1$  is always accepted, then  $n = m - 1$  is also always accepted, and hence,  **$v_m \geq 2$  for  $m \geq 3$**  (the two numbers coincide for  $m = 2$ , so that  $v_2 = 1$ ).
3.  **$v_m$  always is an even number for  $m \geq 3$**  because, according to point 2, the accepted numerators  $n$  always show up in pairs; moreover, **if  $m \geq 3$  is even, then  $n = m/2$  is not accepted** because, for  $m = 2\ell$  (and  $\ell \geq 2$ ), the numerator would be  $n = m/2 = \ell$ , and  $n/m = \ell/2\ell$  would be a reducible fraction.
4. **For  $m \geq 3$ , the sum of an accepted pair always is 1** because we are adding  $k/m$  and  $(m-k)/m = 1 - k/m$ ; as a consequence, **the sum of the irreducible fractions sharing a common denominator  $m$  is  $\sigma_m = v_m/2$**  because there are  $v_m/2$  accepted pairs; looking, moreover, at Table 3, we see that this last result also holds for  $m = 1$  ( $v_1 = 2, \sigma_1 = 1$ ) and  $m = 2$  ( $v_2 = 1, \sigma_2 = 1/2$ ).

### 8. Conclusions (Written by Giovanni M. Cicuta)

Let us summarize the findings of this work. The goal of the work is to define a uniform probability distribution to every rational number  $q \in [0, 1]$ . This does not seem to have practical applications today, but it is an intriguing problem. Indeed, it may only be achieved as a limiting process. The main steps are:

For every ratio of pair of random variables  $N/M$ , with  $n \leq m$ , a uniform distribution for the random variable  $N$  is chosen, for fixed  $m$ . Next, a sequence, indexed by a parameter  $k$  of distributions, is chosen for the random variable  $M$ , the denominator.

The  $k$ -dependent distribution is increasingly more flat as  $k \rightarrow \infty$ .

In terms of these distributions, a  $k$ -dependent probability distribution  $P\{Q = q\}$  is defined in Equation (1), for any  $q = n/m, 0 \leq q \leq 1$ , where the pair  $n, m$  are co-primes. It is defined as *asymptotically equiprobable*. Theorem 1 asserts that all the desired properties for the probability distribution are achieved through this procedure.

Section 6 provides examples of the distributions for the denominator and detailed evaluations.

Section 7 is not probabilistic. The proof of the countability of rational numbers by G. Cantor provides a sequence of all fractions  $n/m$ . Several other sequences were written in the decades following this proof. In all of them, every rational number  $q$  appears several times (an infinite number of times). One would like to define a sequence of all rational numbers  $q \in [0, 1]$ , such that the sequence of fractions  $q = n/m$  would only include the

pair  $n, m$  as co-primes. In Section 7, some properties of the occurrence of repeated fractions are obtained.

If the desired sequence of rational numbers without repetitions could be built, many probability distributions could be proposed for the elements of the sequence.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The author declares no conflict of interest.

## Reference

1. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series and Products*; Academic Press: Burlington, VT, USA, 2007.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.