Sequences of density $\zeta(k) - 1$

William J. Keith

August 9, 2008

1 Introduction

At a recent post-seminar gathering, Herb Wilf casually mentioned to those of us assembled the fact that the quantities $\zeta(k) - 1$ sum to 1, i.e.,

$$\sum_{k=2}^{\infty} \left(\left(\sum_{i=1}^{\infty} \frac{1}{i^k} \right) - 1 \right) = 1 \, .$$

He then declared that when a sequence of nonzero positive numbers sums to 1, the entries of the sequence should well be interpretable as the probabilities of *something*, and asked what it might be in this case. The implied challenge here is that a sequence as interesting as the numbers $\zeta(k) - 1$ should be the probabilities of events equally interesting: preferably rich with structure and easy to describe. One criterion for a really good answer to this challenge would be that the response supplies us with an intuitive feel for the relative sizes of the quantities $\zeta(k) - 1$.

The context of the discussion was number-theoretic, so after some consideration I formalized the challenge thus:

Problem 1. *Partition $\mathbb{N}$ into sets $\{\mathcal{A}_2, \mathcal{A}_3, \dots\}$ such that the asymptotic densities $\lim_{n \to \infty} \frac{1}{n} \# \left(\mathcal{A}_k \cap \{1, 2, \dots, n\} \right) = \zeta(k) - 1$.*

This note provides one such partition. There are, it will be seen, others, but æsthetics guides the choice presented here: the partition can be described in terms of component sets that relate to each other, and we can produce a membership algorithm that can quickly determine which $\mathcal{A}_k$ any number should belong to.

2 The classes $\mathcal{B}_k$

The key both to constructing the classes $\mathcal{A}_k$ and to proving that they have the required densities is to break down the $\zeta(k) - 1$ thus:

$$\begin{array}{llllll}
\zeta(2)-1 & = \frac{1}{2^2} & +\frac{1}{3^2} & +\frac{1}{4^2} & +\frac{1}{5^2} & +\dots \\[4pt]
\zeta(3)-1 & = \frac{1}{2^3} & +\frac{1}{3^3} & +\frac{1}{4^3} & +\frac{1}{5^3} & +\dots \\[4pt]
\zeta(4)-1 & = \frac{1}{2^4} & +\frac{1}{3^4} & +\frac{1}{4^4} & +\frac{1}{5^4} & +\dots \\[4pt]
\zeta(5)-1 & = \frac{1}{2^5} & +\frac{1}{3^5} & +\frac{1}{4^5} & +\frac{1}{5^5} & +\dots \\[4pt]
\ \ \vdots & \ \ \vdots & \ \ \vdots & \ \ \vdots & \ \ \vdots & \\[4pt]
1 & = \frac{1}{2} & +\frac{1}{6} & +\frac{1}{12} & +\frac{1}{20} & +\dots
\end{array} \tag{1}$$

where the entries of the last line are the summations of the (infinite) columns above. The strategy is now to construct classes each of which have asymptotic densities given by the individual entries of the table. As those entries are all unit fractions, our task is much simplified: ask for a set of integers of some peculiar irrational density and intuition would not necessarily leap to service; but ask for one-sixth of the integers, and any residue class mod 6 will do. So we begin in this section with that last row, by producing classes $\mathcal{B}_j$ of densities $\frac{1}{2}, \frac{1}{6}, \frac{1}{12}, \dots$, representing each column. In the next section we break each $\mathcal{B}_j$ apart into sections $\mathcal{B}_{j,k}$ of densities $\frac{1}{j^k}$, $2 \le j, k < \infty$, to produce the individual entries of the table. The finishing stroke is then to collect across rows to form the $\mathcal{A}_k$.

The constructions we use are often much simplified by putting them in terms of the *factorial-base* representation of a number, its expansion as $x = a_1 \cdot 1! + a_2 \cdot 2! + \dots + a_n n!$, $0 \le a_i \le i$). This expansion is unique, as the student unfamiliar with this base system is encouraged to try proving on their own. For more on such expansions, see [3].

The class $\mathcal{B}_2$ must have asymptotic density $\frac{1}{2}$. Let us select it to consist of those $x \equiv 1 \bmod 2$.

Modulo 6, the classes 1, 3, and 5 have been assigned, so for $\mathcal{B}_3$ we select the residue class 2 mod 6.

Modulo 24, we require two residue classes in order that $\mathcal{B}_4$ should have density $\frac{1}{12}$. We select the first two available: those $x \equiv 4$ or 6 mod 24.

Modulo 120, we require six residue classes in order that $\mathcal{B}_5$ should have density $\frac{1}{20}$, and the first six available are 10, 12, 16, 18, 22, and 24.

Notice that we cannot use modulus 20 in the previous step: every residue class mod 20 has already had some subprogressions assigned. We proceed instead through the moduli $j!$. Since for $m < j$, residue classes modulo $m!$ are the same as $\frac{j!}{m!}$ equally spaced residue classes modulo $j!$, all previous assignments can be viewed as assignments modulo $j!$. For each $\mathcal{B}_j$, we select the first $(j-2)!$ residue classes that have not yet been assigned. What are these in general?

Lemma 1. *The assigned residue classes for $\mathcal{B}_m$ are*

$$\left\{ x \equiv 0! + \sum_{j=1}^{m-2} a_j \cdot j! \bmod m!\, , 1 \le a_j \le j \right\}.$$

Proof. In the examples above, by the time we have selected component residue classes for $\mathcal{B}_j$ we have assigned membership in some $\mathcal{B}_i$ to the natural numbers

$1, 2, \ldots, (j-1)!$, as part of residue classes either modulo $j!$ or modulo divisors thereof (previous factorials). Let us proceed by induction, and suppose that this has been the case through $\mathcal{B}_{m-1}$: for each $b \leq m-1$, the numbers $1, 2, \ldots (b-1)!$ have been assigned to some $\mathcal{B}_j$, $j \leq b$, as part of residue classes modulo $j!$.

Given this induction hypothesis, what is the first residue class modulo $m!$ available for assignment to $\mathcal{B}_m$? None of 1 through $(m-2)!$, to begin with: each of those classes were assigned to $\mathcal{B}_{m-1}$ or some previous class. We do know that none of the next $(m-2)!$ residue classes mod $m!$, $(m-2)!+1$ through $(m-2)!+(m-2)!$, are assigned to $\mathcal{B}_{m-1}$. Examine this segment.

Since $(m-2)! \equiv 0 \bmod (m-2)!$, our assumption tells us that the numbers $(m-2)!+1$ through $(m-2)!+(m-3)!$ must have been assigned to $\mathcal{B}_{m-2}$ or earlier classes, as residues modulo $(m-2)!$ or divisors thereof. However, none of $(m-2)!+(m-3)!+1$ through $(m-2)!+(m-3)!+(m-3)!$ were assigned to $\mathcal{B}_{m-2}$, and this segment is wholly contained within those numbers we already knew were not assigned to $\mathcal{B}_{m-1}$. At each stage of the analysis, we have an unbroken segment of assigned classes for the numbers $1, 2, \ldots, (m-2)!+(m-3)!+\cdots+(m-i)!$, and we know that the next $(m-i)!$ numbers are not assigned to $\mathcal{B}_i$ through $\mathcal{B}_{m-1}$. At the last step, we find that the last number assigned is $(m-2)!+(m-3)!+\cdots+2!+1!$, an odd number in class $\mathcal{B}_2$.

The very next number is in the segment we know has not been assigned to $\mathcal{B}_2$ or any previous class, so it is available. Thus the first available residue class modulo $m!$ to assign to $\mathcal{B}_m$ is $(m-2)!+(m-3)!+\cdots+2!+1!+0!$, a construction for which there are unfortunately multiple labels and notations in current usage. The relevant sequence $1, 2, 4, 10, 34, 154, 874, \ldots$ is Sloane's A003422 ([2]) missing a leading zero term, and is there called the *left factorial* and denoted $!n = 0! + \cdots + (n-1)!$. We will use this notation, so that the first unassigned class modulo $m!$ is $!(m-1)$.

Now let us determine the remainder of the assigned classes mod $m!$. At the next-to-last step of our analysis previously, assuming $m \geq 4$ so that such a step exists, we determined that $!(m-1)-2$ must have been an assigned class modulo $3!$, that is, $!(m-1)-2 \equiv 2 \bmod 6$. Then $!(m-1)$ is not such a class and neither is $!(m-1)+2$: these are the two even classes mod 6 that were not assigned to $\mathcal{B}_3$. Nor can they be assigned to $\mathcal{B}_i$ for $i \geq 3$ as part of residue classes modulo multiples of $3!$. Thus, we can select these two as our first two residue classes modulo $m!$ for $\mathcal{B}_m$. If $m \geq 5$, we can ascend backward another step: there we find that $!(m-1)-4$ and $!(m-1)-6$ had to be the assigned residue classes mod $4!$. We already knew $!(m-1)$ and $!(m-1)+2$ were not in those classes, and here we find that they can be joined by $!(m-1)+6$ or $+12$, and $!(m-1)+2+6$ or $+12$. These are the six residue classes modulo 24 that were left after we assigned two of them for $\mathcal{B}_4$.

At the i-th step of the ascent, we find that to all the entries we previously picked, we can add as many as $i-1$ multiples of $i!$. Thinking about it in the factorial-base representation, using $a_1 a_2 a_3 \ldots$ to denote $a_1 \cdot 1! + a_2 \cdot 2! + a_3 \cdot 3! + \ldots$, the assigned classes will be $02111 \cdots + 0 b_2 b_3 b_4 b_5 \ldots$, with $0 \leq b_i \leq i-1$. And these are exactly the numbers claimed in the Lemma. $\square$

2.1 A Fast Membership Algorithm

Having assigned the whole numbers to their various $\mathcal{B}_m$, we would like a means of determining which $\mathcal{B}_m$ a given x belongs to without constructing every $\mathcal{B}_m$ until we assign x. As it turns out, systematically selecting the first available classes at each step helps us write a computationally simple algorithm to do so.

Write x in the factorial-base representation, $x = x_1 \cdot 1! + x_2 \cdot 2! + x_3 \cdot 3! + \cdots + x_j \cdot j!$, $0 \leq x_i \leq i$. We know that if the least positive residue of x modulo any $n!$ is less than $(n-1)!$, x must have been assigned to one of the $\mathcal{B}_{j \leq n}$, because from our proof above, the assigned residue classes for the $\mathcal{B}_{j \leq n}$ were an unbroken string from 1 to $(n-1)!$.

Now, a "carry" in factorial-base addition occurs when the sum of the coefficients on the two summands on $i!$ is at least $i+1$, since $(i+1)i! = 1 \cdot (i+1)!$. But the Lemma tells us that the residues assigned to $\mathcal{B}_m$ are of the form $02111\cdots + 0b_2b_3b_4b_5\ldots$, with $0 \leq b_i \leq i-1$. So in order for a 0 to ever occur at the $i!$ place of the factorial-base expansion of a residue assigned to $\mathcal{B}_m$, when $1 < i \leq m-1$, we need a very specific summand. We need $b_2 = 1$, so that we carry from 2!, leaving a 0. We then need $b_3 = 2$, so that the sum of coefficients on 3! is 1 from $!(m-1)$, plus 2 from b_3, plus a carried 1, making 4, so we carry 1 and leave a 0 again. This string has to continue up to the $i!$ term where we desire a 0.

Thus we have one of two cases that will diagnose where $!(m-1)$ ends for a given x. We have that the least positive residue of x modulo $m!$ is $\leq (m-1)!$. Possibly $x \equiv (m-1)!$ modulo $m!$ exactly, so the string of x_i starts with a possibly empty string of 0s that terminates with a 1 in position x_{m-1}: we added the largest possible value for every b_i, and carried at every step of the addition. If we did not carry, then there is some smallest m such that $x_{m-1} = 0$ but no previous $x_i = 0$, except for possibly an initial string that does *not* terminate with a 1. Write out x in the factorial-base representation, check for the first behavior, and if it doesn't happen find the first 0. The index where whichever of these occurs, occurs, is $m-1$.

3 The Classes $\mathcal{A}_k$

Now that we have identified the classes $\mathcal{B}_m$, there remain the tasks of breaking them up into subclasses to give us the individual entries of Table 1, and collecting the corresponding parts "horizontally" to form the $\mathcal{A}_k$.

Each of the columns' individual entries decrease in geometric progression. The class $\mathcal{B}_2$ has density $\frac{1}{2} = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots\right)$. The class $\mathcal{B}_3$ has density $\frac{1}{6} = \frac{1}{6}\left(\frac{2}{3} + \frac{2}{9} + \frac{2}{27} + \ldots\right)$. The class $\mathcal{B}_4$ has density $\frac{1}{12} = \frac{1}{12}\left(\frac{3}{4} + \frac{3}{16} + \ldots\right)$, etc.

These geometric series give us the subclasses we need. We assign the earliest fraction $\frac{m-1}{m}$ of $\mathcal{B}_m$ to the first subset, the next fraction $\frac{1}{m}\frac{(m-1)}{m}$ to the next set, and so forth. This can be done by a one-step digit test. For an example from familiar territory, to obtain $\frac{9}{10}$ of the integers, take those ending in 1 through 9.

Then to obtain $\frac{9}{100}$ of the integers by choosing $\frac{9}{10}$ of the remaining tenth, take those that end in 0 but have one of 1 through 9 in the tens place.

Since $\mathcal{B}_m$ is made up of residue classes modulo $m!$, rewrite x as $x = x_0 + m! \cdot \left(b_0 m^0 + b_1 m^1 + b_2 m^2 + \dots\right)$, $0 \leq b_i \leq m - 1$. If n is the smallest number such that $b_n \neq m - 1$ assign x to the set $\mathcal{B}_{m,n+2}$. In table 1, x is part of the set with asymptotic density given by the entry in column m, row $n + 1$, which represents an entry in the sum for $\mathcal{A}_{n+2}$.

Thus, for example, we break up $\mathcal{B}_2$ as follows: $\mathcal{B}_{2,2} = \{x \equiv 1 \bmod 4 (= 2 \cdot 2!)\}$, with density $\frac{1}{4}$; $\mathcal{B}_{2,3} = \{x \equiv 3 \bmod 8 (= 4 \cdot 2!)\}$, with density $\frac{1}{8}$; $\mathcal{B}_{2,4} = \{x \equiv 7 \bmod 16 (= 4 \cdot 2!)\}$, etc. We break up $\mathcal{B}_3$ into fractions of size $\frac{2}{3}, \frac{2}{9}$, etc.: $\mathcal{B}_{3,2} = \{x \equiv 2 \text{ or } 8 \bmod 18 (= 3 \cdot 3!)\}$, $\mathcal{B}_{3,3} = \{x \equiv 14 \text{ or } 32 \bmod 54 (= 9 \cdot 3!)\}$, etc.

We now sum up by collecting these subsets across rows: $\mathcal{A}_k = \bigcup_m \mathcal{B}_{m,k}$. Since the $\mathcal{B}_{m,k}$ are disjoint and the series of partial sums of their densities converges absolutely to $\zeta(k) - 1$, $\mathcal{A}_k$ has exactly the required asymptotic density.

The algorithm to determine which class $\mathcal{A}_k$ a given x belongs to is thus:

Algorithm. *Write $x = a_1 \cdot 1! + a_2 \cdot 2! + a_3 \cdot 3! + \dots$, $0 \leq a_i \leq i$.*

Let j be the smallest index such that $a_j \neq 0$. If $a_j = 1$, then $m = j + 1$. If $a_j \neq 1$, then m is the smallest number such that $a_i \neq 0$ for $j \leq i < m - 1$, and $a_{m-1} = 0$. Then $x \in \mathcal{B}_m$.

Rewrite x as $x = x_0 + m! \cdot (b_0 m^0 + b_1 m^1 + b_2 m^2 + \dots)$, $0 \leq b_i < m$, $0 < x_0 < m!$. Let n be the smallest index such that $b_n \neq m - 1$. Then $k = n + 2$, and $x \in \mathcal{A}_k$.

4   Summary and Speculation

The first few of the sets $\mathcal{B}_k$ and $\mathcal{A}_k$ are listed below.

$\mathcal{B}_2 = \{1, 3, 5, 7, 9, 11, 13, 15, \dots\}$.

$\mathcal{B}_3 = \{2, 8, 14, 20, 26, 32, \dots\}$.

$\mathcal{B}_4 = \{4, 6, \quad 28, 30, \quad 52, 54, \quad 76, 78, \dots\}$.

$\mathcal{B}_5 = \{10, 12, 16, 18, 22, 24, \quad 130, 132, 136, 138, 142, 144, \dots\}$.

$\mathcal{B}_6 = \{34, 36, 40, 42, 46, 48, 58, 60, 64, 66, 70, 72, \dots, 120, \dots\}$.

$\mathcal{A}_2 = \{1, 2, 4, 5, 6, 8, 9, 10, 12, 13, 16, 17, 18, 20, 21, 22, 24, 25, 26, \dots\}$.

$\mathcal{A}_3 = \{3, 11, 14, 19, 27, 32, 35, 43, 51, 59, 67, 68, 75, 76, 78, 83, 86, \dots\}$.

$\mathcal{A}_4 = \{7, 23, 39, 50, 55, 71, 87, 103, 104, 119, 135, 151, 167, 183, 199, \dots\}$.

$\mathcal{A}_5 = \{15, 47, 79, 111, 143, 158, 175, 207, 239, 271, 303, 320, 335, \dots\}$.

More terms, for the series up to $\mathcal{A}_{10}$, have been submitted to Sloane's database [2] and should be available by the time of publication of this note.

A bit more illuminating may be the following description of the classes:

$\mathcal{B}_2 = \{x \equiv 1 \bmod 2\}$.

$\mathcal{B}_3 = \{x \equiv 2 \bmod 6\}$.

$\mathcal{B}_4 = \{x \equiv 4, 6 \bmod 24\}$.

$\mathcal{B}_5 = \{x \equiv 10, 12, 16, 18, 22, 24 \bmod 120\}$.

$\mathcal{B}_6 = \{x \equiv 34, 36, 40, 42, 46, 48, 58, 60, 64, 66, 70, 72, \dots, 120 \bmod 720\}$.

$\mathcal{A}_2 = \{x \equiv 1 \bmod 4; \ 2, 8 \bmod 18; \ 4, 6, 28, 30, 52, 54 \bmod 96 \dots\}$.

$\mathcal{A}_3 = \{x \equiv 3 \bmod 8; \ 14, 32 \bmod 54; \ 76, 78, 172, 174, 268, 270 \bmod 384 \dots \}$.

$\mathcal{A}_4 = \{x \equiv 7 \bmod 16; \ 50, 104 \bmod 162; \ 364, 366, \dots, 1134 \bmod 1536 \dots \}$.

Some variations of this construction could be explored. Building $\mathcal{B}_j$, we had to make a series of choices. We chose the odd numbers to form $\mathcal{B}_2$, but the even numbers would have worked just as well. We chose the residue class $x \equiv 2 \bmod 6$ for the class $\mathcal{B}_3$, but could just as easily have chosen $x \equiv 4$ or 6. Choosing the first available residue class for each factorial modulus gave us classes $\mathcal{B}_m$ that could be determined with a short algorithm, providing a tidy answer to the original problem. However, other choices lead to answers with different features.

For example, suppose that at each step we choose the *last* available residue class modulo $j!$ to construct $\mathcal{B}_j$. Choose those $x \equiv 2 \bmod 2$ for $\mathcal{B}_2$. Of the three available odd residue classes mod 6, choose $x \equiv 5 \bmod 6$ for $\mathcal{B}_3$. Use the classes $x \equiv 19, 21 \bmod 24$ for $\mathcal{B}_4$. With such a decision procedure, some numbers are never assigned to any $\mathcal{B}_j$ at all!

The "missed set" is $1, 3, 7, 9, 13, 15, 25, 27, 31, 33, \dots = 1 + \sum a_i i!, 0 \le a_i < i$. All of the $\mathcal{B}_j$ will still have the same densities, summing to 1, so the missed set is "small" in that it is of asymptotic density 0. In fact, given any arithmetic progression $X \bmod Y$, some subprogression will be assigned to a $\mathcal{B}_j$. (Show it!) On the other hand, at any given point the missed set may be rather large for some purposes: $1/n$ of the numbers smaller than $n!$ are permanently unassigned.

These two assignment procedures are in some sense on opposite poles of an entire ensemble of possible procedures. An interested reader might burrow a layer deeper than we have: assign some straightforward process for describing and choosing assignment procedures to construct each $\mathcal{B}_j$ from residue classes mod $j!$ and examine the resulting ensemble of all possible constructions. Does it possess any striking structural features?

An early stab at this problem involved the powerfree numbers. Squarefree numbers are those whose prime factors are all distinct: they have density $\frac{1}{\zeta(2)}$ in the whole numbers. Cubefree numbers have factorizations in which no prime factor is repeated more than twice, so squarefree numbers are also cubefree. Cubefree numbers are of density $\frac{1}{\zeta(3)}$, and those which are cubefree but not squarefree are of density $\frac{1}{\zeta(3)} - \frac{1}{\zeta(2)}$. Those that are 4th-powerfree but not cubefree are of density $\frac{1}{\zeta(4)} - \frac{1}{\zeta(3)}$, and so on. Every whole number is kth-powerfree but not $(k+1)$st-powerfree for some k, the largest exponent in its prime factorization.

These densities seemed to suggest the possibility of defining a simple probability distribution on $\mathbb{N}$ that assigned a total probability of $\zeta(k) - 1$ to the event that a random integer variable would be kth-powerfree but not $(k+1)$st-powerfree. It is easy to provide such a distribution by brute force – say, giving the nth number which is kth-powerfree but not $(k+1)$st-powerfree a probability of $\frac{\zeta(k)-1}{\zeta(k)} \frac{1}{2^n}$. But this does not seem to be a particularly illuminating illustration of any relations between the values $\zeta(k)$. A distribution based on the factorization of x would seem much more natural; can a simple one be produced?

In closing, I would like to mention a personal recollection. In general, the density of $\mathcal{B}_j$ is $\frac{1}{j(j-1)}$. Seeing it written on the wall of my office, I was reminded

6

of the first place I ever encountered Leibniz' summation of the triangular series: William Dunham's delightful popular-mathematics text, *Journey Through Genius* [1]. I read this book in high school, and it motivated in considerable part my decision to pursue mathematics in college. The present note gives me an opportunity to thank Mr. Dunham sincerely for the service.

References

[1] Dunham, William. *Journey Through Genius: The Great Theorems of Mathematics.* John Wiley and Sons, 1990. ISBN 0-471-50030-5.

[2] Sloane, N.J.A., 2008. The On-Line Encyclopedia of Integer Sequences, published electronically at www.research.att.com/ njas/sequences/

[3] Barwell, Brian R. Factorian Numbers. *Journal of Recreational Mathematics*, 7:63, 1974.