Examples of integer sequence density calculations

There are many ways that the density of an integer sequence has been considered.

Perhaps the most natural is the *natural density*. For a sequence of positive integers S, define $A(x)$ to be the counting function for S, i.e.,

$$A(x) = |\{n \in S : 1 \le n \le x\}|.$$

So $A(x)$ is the number of terms in S that are less than or equal to x. Then, the natural density of S is defined to be

$$d(S) = \lim_{x \to \infty} \frac{A(x)}{x}$$

provided the limit exists.

We see that if S consists of all positive integers, then $A(x) = x$, and so $d(S) = 1$.

If S consists of all odd positive integers, then $A(x) = \lfloor \frac{x+1}{2} \rfloor$, so

$$\frac{x+1}{2} - 1 \le A(x) \le \frac{x+1}{2}$$

and so

$$\frac{1}{2} - \frac{1}{2x} \le \frac{A(x)}{x} \le \frac{1}{2} + \frac{1}{2x}.$$

Hence, $\lim_{x \to \infty} \frac{A(x)}{x} = \frac{1}{2}$, and so $d(S) = \frac{1}{2}$ for the odd positive integers.

If we let m be a positive integer, and B a set of residue classes modulo m, and let S be the set of all positive integers that fall into an element of B, then, with an argument similar to the one above,

$$d(S) = \frac{|B|}{m}.$$

So for example, the set of integers congruent to 1, 4 or 7 modulo 8 has natural density $\frac{3}{8}$.

Can the limit not exist?

The definition of the natural density requires the existence of a limit. It turns out that there are sequences for which the limit does not exist, and these sequences do not have a natural density.

Consider the set of integers S which have an odd number of digits. Consider some values of $\frac{A(x)}{x}$. We have $\frac{A(9)}{9} = 1$, $\frac{A(99)}{99} = \frac{9}{99} = \frac{1}{11}$, $\frac{A(999)}{999} = \frac{909}{999} = \frac{101}{111}$, and $\frac{A(9999)}{9999} = \frac{909}{9999} = \frac{1}{11}$. One can show that, if $x = 10^k - 1$ and k is even, then $\frac{A(x)}{x} = \frac{1}{11}$, while if k is odd, then $\frac{A(x)}{x} > \frac{9}{10}$. As a result of this behavior, we can see that $A(x)/x$ does not approach a limit as x approaches infinity, and so this set does not have a natural density.

It is not hard to construct many sequences which fail to have natural densities.

Example: partitioning the sequence

Consider the sequence of numbers, S, which have a binary representation that ends in an even number of zeros. This is A003159 in the OEIS, and begins $1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, \ldots$.

Let's work out what the natural density of this sequence is.

A very helpful approach here is to partition the sequence into disjoint subsequences and work out the density of each. A nice property of the natural density is that it is **additive**: if disjoint sequences S_1 and S_2 have natural densities, then the natural density of their union exists and we have

$$d(S_1 \cup S_2) = d(S_1) + d(S_2).$$

(The proof of this fact makes a nice exercise.)

To partition S, we can note that it consists of numbers with binary expansions ending in exactly 0 zeros, exactly 2 zeros, exactly 4 zeros, etc.

We can characterize precisely each type.

The set of numbers with binary expansions ending in exactly 0 zeros is the set of odd numbers, i.e. the numbers congruent to 1 mod 2.

The set of numbers with binary expansions ending in exactly 2 zeros is the set of all numbers congruent to 4 mod 8.

The set of numbers with binary expansions ending in exactly 4 zeros is the set of all numbers congruent to 16 mod 32.

Generally, the set of numbers with binary expansion ending in exactly $2k$ zeros is the set of all numbers congruent to 2^{2k} mod 2^{2k+1}.

As mentioned earlier, the density of such a set is

$$\frac{1}{2^{2k+1}}$$

and so the density of the set S of number with binary expansions ending in an even number of zeros is

$$d(S) = \sum_{k=0}^{\infty} \frac{1}{2^{2k+1}} = \frac{2}{3}.$$

Example: probability heuristic and Dirichlet density

Let's consider the sequence S of all numbers whose prime factorization consists only of odd exponents. S contains all primes, all odd powers of primes (e.g., 3^3, 7^5, etc.), and all products of relatively odd powers of primes, e.g., $2^5 3^3 11^9$.

This is sequence A002035 at the OEIS.

The sequence begins, $2, 3, 5, 6, 7, 8, 10, 11, 13, 14, 15, 17, 19, 21, 22, \dots$. If we count terms, we find $A(100) = 70$, $A(10000) = 7055$ and $A(10^6) = 704464$, which experimentally suggests the sequence might have a natural density around 0.7.

Here's a way we can get a good handle on the density. We begin by noting that, if an integer is *not* in this sequence, then it must have an even exponent somewhere in its prime factorization. That is, it must be exactly divisible by an even power of a prime.

For any prime p and positive integer k, the density of integers that are exactly divisible by p^k is

$$\frac{1}{p^k} - \frac{1}{p^{k+1}}.$$

This comes from the additivity property of the natural density mentioned earlier. Hence, the density of positive integers that are exactly divisible by an even power of p is

$$\frac{1}{p^2} - \frac{1}{p^3} + \frac{1}{p^4} - \frac{1}{p^5} + \frac{1}{p^6} - \frac{1}{p^7} + \cdots$$

and so the density of numbers that are not exactly divisible by an even power of p is

$$1 - \frac{1}{p^2} + \frac{1}{p^3} - \frac{1}{p^4} + \frac{1}{p^5} - \frac{1}{p^6} + \frac{1}{p^7} - \cdots = \frac{p^2 + p - 1}{p^2 + p}.$$

For example, the set of numbers not exactly divisible by an even power of 2 has density $\frac{5}{6}$ and the set of numbers not exactly divisible by an even power of 3 has density $\frac{11}{12}$.

Now, heuristically, treating these densities as probabilities we can calculate the density of the full sequence S. For an integer n to be in S, it must not by exactly divisible by an even power of 2, and it must not be exactly divisible by an even power of 3, and it must not be divisible by an even power of 5, etc.

Hence, the probability that the number is not divisible by an even power of any prime is

$$\prod_p \frac{p^2 + p - 1}{p^2 + p} = \prod_p \left(1 - \frac{1}{p(p+1)}\right).$$

where the product is over all primes. This is the density of the set S.

A more formal argument can be made with Dirichlet series. This begins by noting that, for any complex z,

$$L(z) = \sum_{n \in S} n^{-z} = \prod_p \left(1 + \frac{1}{p^z} + \frac{1}{p^{3z}} + \frac{1}{p^{5z}} + \cdots\right) = \prod_p \frac{1 + p^{-z} - p^{-2z}}{1 - p^{-2z}}.$$

Then the *Dirichlet density* of S, $\mu(S)$ is given by

$$\mu(S) = \text{Res}_{z=1} L(s) = \lim_{z \to 1^+} \frac{L(z)}{\zeta(z)}$$

$$= \lim_{z \to 1} \prod_p \frac{1 - 2p^{-2z} + p^{-3z}}{1 - p^{-2z}}$$

$$= \prod_p \frac{1 - 2p^{-2} + p^{-3}}{1 - p^{-2}} = \prod_p \left(1 - \frac{1}{p(p+1)}\right).$$

which is the same as we worked out heuristically above.

If we calculate this product through the millionth prime, the result is

$$0.704442203598081\ldots$$

and by bounding the tail we can conclude that the true density $d(S)$ is in the range

$$0.704441 < d(S) < 0.704442.$$