INTRODUCING SOFIA NUMBERS
AND NEW CONJECTURES ON PRIMES

(Umberto Bartocci - Alessandro Miotto)

\- - - - -

Abstract - Given any natural number n, we consider the subset C(n) of all natural numbers which could be written in decimal basis as a string of length n (n-numbers). We are looking for subsets of C(n) which appear interesting in connection with n-prime numbers and twin n-prime numbers too. Those numbers consisting of only the digits 1 and 7 (that we call **Sofia numbers**) appear to be quite promising in this context, and their study suggests some natural conjecture.

\- - - - -

For any natural number $n \in N$, let us define the subset $C(n) \subset N$ of all natural numbers which could be written in decimal basis as a string of length n (n-numbers), and then its subset $S(n) \subset C(n)$ of those n-numbers which have only the ciphers 1 and 7. We shall call these numbers **Sofia numbers** (or n-Sofia numbers for a given length n), in honour of the little niece of the oldest of the authors.

With obvious meaning of the symbols (including the here unusual $\oplus$):

$S(1) = \{1, 7\}$
$S(2) = 1 \oplus S(1) \cup 7 \oplus S(1) = \{11, 17\} \cup \{71, 77\} = \{11, 17, 71, 77\}$
$S(n) = 1 \oplus S(n-1) \cup 7 \oplus S(n-1) = \{1111...1, 1...17, 1...171, ... , 7777...7\}$.

It is clear that $|S(n)| = 2^n$, and that, for $\forall$ x, y $\in$ S(n): $x \equiv y \bmod 6$. In words, all n-Sofia numbers have the same residual class mod 6, and these residual classes are described by the following simple scheme (all congruences mod 6):
$S(1) \equiv 1$, $S(2) \equiv -1$, $S(3) \equiv 3$, $S(4) \equiv 1$, $S(5) \equiv -1$, $S(6) \equiv 3$, ...
That is to say, the residual class mod 6 of all n-Sofia numbers is the "same" residual class of n mod 3 in the cases 1 or -1, and it is 3 when this residual class of n mod 3 is 0.

n-Sofia numbers appear to be interesting in connection with the research of n-prime numbers, namely $PC(n)=P \cap C(n)$ (P the set of all prime numbers; remember that indeed all prime numbers greater than 3 are either congruent to 1 mod 6 or to -1 mod 6), and in such a particular context with the research of **twin prime numbers**, prime numbers p such that either p+2 or p-2 is a prime number too. As a matter of fact, n-Sofia numbers (**when n is not a multiple of 3**, case in which there are not n-Sofia prime numbers) appear to have a propensity to primality greater than the analogous propensity of the other n-numbers. We shall try to give empirical evidence of this conjecture.

For any finite subset $X \subset N$, let us introduce the arithmetical functions:
$\pi(X) = |P \cap X|$
$\pi_2(X)$ = number of twin primes in $P \cap X$
and then try to compare $\pi(C(n))$ with $\pi(S(n))$, or $\pi_2(C(n))$ with $\pi_2(S(n))$, or better the corresponding **probabilities**:
$\pi(C(n))/|C(n)|$ with $\pi(S(n))/|S(n)|$
$\pi_2(C(n))/\pi|C(n)|$ with $\pi_2(S(n))/\pi(S(n))$

Note - It is obvious that the prime numbers "companions" of twin n-Sofia prime numbers are not n-Sofia numbers, which is not the case of course for n-numbers in general.

The aforesaid probabilities $\pi(S(n))/|S(n)|$ and $\pi_2(S(n))/ \pi(S(n))$, for n up to 28, can be deduced by the following numerical tables, in which $\sigma(n)=\pi(S(n))$, and $\gamma(n)=\pi_2(S(n))$ (remember that we discard values of n which are multiples of 3):

Table I
Number of ciphers | number of primes | number of twin primes |
% $\sigma(n)/2^n$ | % $\gamma(n)/\sigma(n)$
n=1 | σ=1 | γ=1 | 50.00% | 100.00%
n=2 | σ=3 | γ=3 | 75.00% | 100.00%
n=4 | σ=5 | γ=0 | 31.25% | 0.00%
n=5 | σ=11 | γ=6 | 34.38% | 54.55%
n=7 | σ=31 | γ=1 | 24.22% | 3.23%
n=8 | σ=43 | γ=11 | 16.80% | 25.58%
n=10 | σ=138 | γ=8 | 13.48% | 5.80%
n=11 | σ=358 | γ=67 | 17.48% | 18.72%
n=13 | σ=1044 | γ=52 | 12.74% | 4.98%
n=14 | σ=1781 | γ=244 | 10.87% | 13.70%
n=16 | σ=5995 | γ=322 | 9.15% | 5.37%
n=17 | σ=12743 | γ=1436 | 9.72% | 11.27%
n=19 | σ=45534 | γ=1623 | 8.68% | 3.56%
n=20 | σ=78005 | γ=6645 | 7.44% | 8.52%
n=22 | σ=297781 | γ=11242 | 7.10% | 3.78%
n=23 | σ=595835 | γ=46786 | 7.10% | 7.85%

n=25 | σ=2200996 | γ=62183 | 6.56% | 2.83%
n=26 | σ=4062195 | γ=254560 | 6.05% | 6.27%
n=28 | σ=14691048 | γ=434646 | 5.47% | 2.96%

Maintaining the hypothesis that n is not a multiple of 3, the first easy conjectures suggested by Table I are:

A - σ(n) is always a positive number, and it is strictly increasing with n;

B - γ(n) is a positive number for infinitely many values of n (this is a kind of **Restricted Twin Prime Conjecture**).

We could strengthen conjecture A in the following way:

C - π(S(n))/|S(n)|= σ(n)/2^n is always greater than π(C(n))/|C(n)|, that is to say, the probability for primality of n-Sofia numbers is greater than the corresponding probability for a general n-number.

By means of the asymptotic estimate of π(C(n)) given by the **Prime Number Theorem**, namely π(n) ~ n/log(n), and from the following nice upper bound that one could deduce from the previous estimate for the prime-counting function π(n):

π(C(n))/|C(n)| < 1/2n

one could reformulate C as σ(n)/2^n > 1/2n, namely:

C* - σ(n) > 2^n/2n.

This is a rather elegant inequality which is of course confirmed by Table I (for each n≥2, for n=1 the strict inequality does not hold).

Question - When C holds for some n=3k+1 and n=3k+2, could this circumstance have something to do with the fact that σ(3k)=0? (that is to say, a kind of "balance" between some expected values should be researched in triplets, and not just in a single S(n), or in pairs of them?)

An exact comparison between π(C(n))/|C(n)| and π(S(n))/|S(n)| could be obtained by means of an analogous (but more difficult to get) numerical table (for n again up to 28, and n>3):

Table II
Number of ciphers | number of primes | number of twin primes |

% $\pi(C(n))/|C(n)|$ | % $\pi_2(C(n))/\pi(C(n))$
n=4 | 1061 | 340 | 11.79% | 3.78%
n=5 | 8363 | 2038 | 9.29% | 2.26%
n=7 | 586081 | 101622 | 6.51% | 1.13%
n=8 | 5096876 | 762664 | 5.66% | 0.85%
n=10 | 404204977 | 47976346 | 4.49% | 0.53%
n=11 | 3663002302 | 393926738 | 4.07% | 0.44%
n=13 | 308457624821 | 27928159304 | 3.43% | 0.31%
n=14 | 2858876213963 | 239891313586 | 3.18% | 0.27%
n=16 | 249393770611256 | 18253972909988 | 2.77% | 0.20%
n=17 | 2344318816620308 | 161289287311722 | 2.60% | 0.18%
n=19 | 209317712988603747 | *12556192074351232* | 2.33% | *0.14%*
n=20 | 1986761935284574233 | *112952263025822814* | 2.21% | *0.13%*
n=22 | 180340017203297174362 | *9282733296595699368* | 2.00% | *0.10%*
n=23 | 1723853104917488062633 | *84724368057308771750* | 1.92% | *0.09%*
n=25 | 158410709631794568543814 | *7140435342710208488148* | 1.76% | *0.08%*
n=26 | 1522400441473293371915923 | *65892536551604946384654* | 1.69% | *0.07%*
n=28 | 141236808849131729966312199 | *5662397225116080090884516* | 1.57% | *0.06%*

(values in red italics are meant to be just approximations).

Conjectures concerning comparisons between the number of twin n-primes and the corresponding number of twin Sofia n-primes are harder to be stated, in absence of good estimates for $\pi_2(C(n))$, and worst for $\pi_2(S(n))$, but we could at least say that one can find interesting the empirical fact that the values given in the last column of Table I are in general greater than the values given in the last column of Table II, which allows us to presume that the probability for a n-Sofia prime number to be a twin is greater than the corresponding probability for a general prime n-number. Of course, it appears well possible to find others criteria for the definition of subsets of C(n) with the same behaviour of S(n) with respect to PC(n)= $P \cap C(n)$, bit this is another question, and perhaps another paper.

Sofia numbers include the so-called **repunits** 11111...1 (repeated units), introduced by Albert Beiler in his 1964 book *Recreations in the Theory of Numbers*. We shall consequently call **quasi-repunits** the Sofia numbers as before but for only one cipher equal to 7, namely, 17, 71, 711. 171, 117, 7111, etc.. Either for repunits or quasi-repunits, it is quite a problem just to establish their possible primality. For instance, only a few repunits are today known to be primes, namely for the lengths 2, 19, 23 (all these lengths are prime numbers too, not a coincidence, but a theorem which holds for all prime repunits), and then 317, 1031, 49081. Larger probable primes have been identified but not proven, that is 86453, 109297, 270343, 5794777 and 8177207 (https://mathworld.wolfram.com/RepunitPrime.html). On the other side, one could ask for instance how many primes appear in the sequence of "special" quasi-repunits 17, 117, 1117, ... , and we discover that a general answer is still not known. Are indeed primes 17 (a twin), 1117, 11117 (another twin), 11111117 (again a twin), 22 times the cipher 1 + one last 7, and furthermore the special quasi-repunits corresponding to the lengths 29, 40, 131, 136, 215. As far as general quasi-repunits is concerning, we have found "long" quasi-repunits which are primes and twin, such [1x411][7][1x1663] (length 2075), which must be read as: 411 times the cipher 1, then one isolated 7, and then again 1663 times the cipher 1. We have found other quasi-repunits which are twin primes, for instance [1x47][7][1x10] (length 58) and [1x269][7][1x179] (length 449). Could all this be a clue for a possible **"even more restricted" Twin Prime Conjecture**? That is to say, restricted not just to Sofia numbers, but to quasi-repunits. It appears instead difficult to propose this same conjecture for special quasi-repunits, since up to the length 1500 do not appear other twins except the ones mentioned above.

In conclusion, one should admit that, even if modern computations are obviously unable to give arithmetical "proofs" when the number of cases is infinite, they can somewhere lead to glimpse otherwise unknown and unexpected phenomenology.

[See also: U. Bartocci-R.V. Macrì. "Numeri primi e numeri di Sofia", *Rudi Mathematici*, N. 321, 2025
https://www.rudimathematici.com/archivio/archiviodb.php]

Umberto Bartocci - bube231##yahoo.it
Alessandro Miotto - amiotto2##gmail.com

Perugia, November 2025