ON INTEGER SEQUENCES WITH MUTUAL k–RESIDUES

SEPPO MUSTONEN

We define an integer sequence $A(k)$

$$a_{k,1},\ a_{k,2},\ \ldots,\ a_{k,n},\ldots$$

with mutual $k-$residues as

$$a_{k,1} = k+1, \quad a_{k,n} = \min\{m \mid m > a_{k,n-1},\ \mathrm{mod}(m, a_{k,i}) \geq k,\ i = 1, 2, \ldots, n-1\}$$

where $\mathrm{mod}(a, b)$ denotes the common residue of $a \pmod b$.

$k = 0$ gives simply natural numbers

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \ldots \quad \text{(Sloane's A000027) [3]}$$

and for $k = 1$ we obtain all prime numbers

$$2, 3, 5, 7, 11, 13, 17, 19, 23 \ldots \quad \text{(Sloane's A000040).}$$

Hence sequences $A(k)$ for $k = 2, 3, \ldots$ can be considered as logical derivatives of natural and prime numbers.

The first 100 numbers of $A(2)$ are

3 5 8 14 23 38 44 53 59 62 68 74 83 122 134 143 158 164 173 179 188 194 203 218 227 242 263 278 284 293 302 314 338 362 374 383 398 404 422 428 443 452 458 467 479 482 503 509 524 539 542 548 554 563 578 614 623 638 653 662 674 698 707 719 734 758 764 773 779 788 794 803 818 839 842 863 878 893 899 923 932 947 974 983 998 1028 1034 1043 1067 1094 1118 1124 1133 1139 1142 1154 1187 1199 1202 1214

For example, in this sequence the fifth term is 23 since

$$\mathrm{mod}(15, 3) = 0,\ \mathrm{mod}(16, 2) = 1,\ \mathrm{mod}(17, 8) = 1,\ \mathrm{mod}(18, 3) = 0,\ \mathrm{mod}(19, 3) = 1,$$
$$\mathrm{mod}(20, 5) = 0,\ \mathrm{mod}(21, 3) = 0,\ \mathrm{mod}(22, 3) = 1$$

but

$$\mathrm{mod}(23, 3) = 2,\ \mathrm{mod}(23, 5) = 3,\ \mathrm{mod}(23, 8) = 7,\ \mathrm{mod}(23, 14) = 9.$$

The first terms of sequences $A(k),\ k = 2, 3, \ldots, 8$ are

2	3	5	8	14	23	38	44	53	59	62	68	74	83	122	134
3	4	7	11	19	27	31	47	75	87	103	131	139	159	179	195
4	5	9	14	24	34	79	89	94	124	134	149	214	229	259	304
5	6	11	17	29	41	65	95	107	161	185	227	251	269	281	305
6	7	13	20	34	48	76	90	111	167	188	216	258	279	349	370
7	8	15	23	39	55	87	103	127	191	247	295	343	359	367	399
8	9	17	26	44	62	98	116	152	332	386	404	539	557	638	674

None of these sequences appear in Sloane's Encyclopedia for the time being (18 Aug 2005).

I have made a Maple procedure for computing $A(k)$ numbers listed below:

Date: August 18, 2005.

Key words and phrases. integer sequences, residues, primes, Survo.

```
----------------------------------------------------------------------
res_seq:=proc(a::array(1,nonnegint),k,n::nonnegint)
local i,j,m,f;
a[1]:=k+1;
for i from 2 to n do
   m:=a[i-1]+1; f:=1;
   while f= 1do
     j:=1;
     while j<i and irem(m,a[j])>=k do j:=j+1; od;
     if j=i then a[i]:=m; f:=0;
     else m:=m+1; fi;
     od;
od;
end;
# Computing 60 first terms of A(2)
a:=array(1..60,[]);
res_seq(a,2,60);
print(a);
----------------------------------------------------------------------
```

For a faster execution I have also made the same thing in C as a SURVO MM module RES_SEQ ([4]).

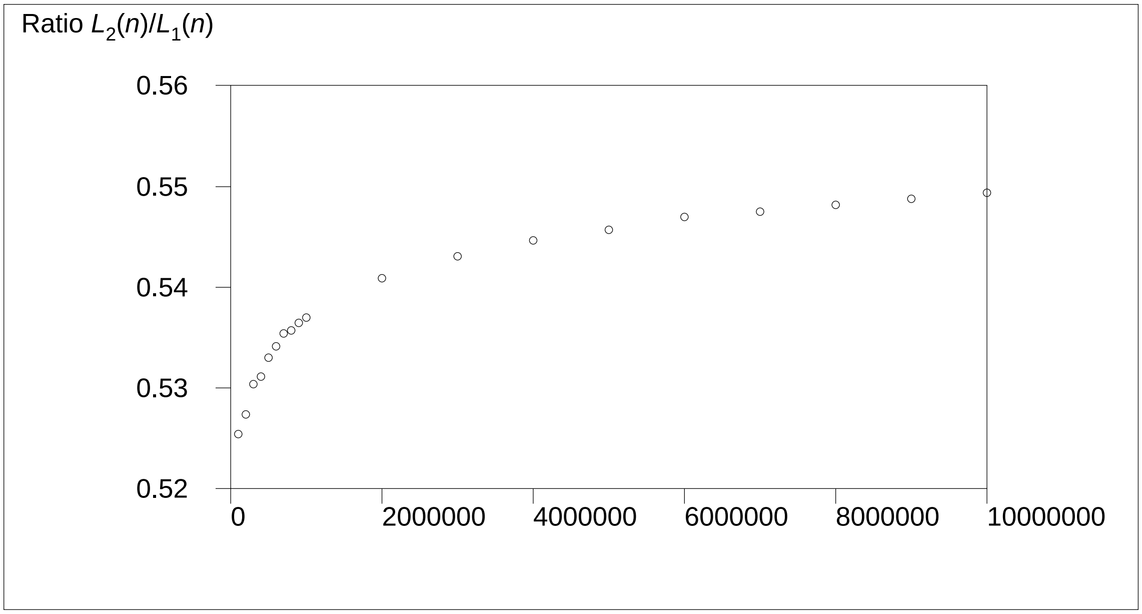
It seems natural to expect that these sequences are related to prime numbers in their large scale behaviour.

By computing $A(2)-$numbers for $n = 1, 2, \ldots, 10^7$ by RES_SEQ the following summary is obtained:

$L_1(n)$ is the number of primes below n. $L_2(n)$ is the number of $A(2)-$numbers below n.

n	$L_2(n)$	$L_1(n)$	ratio
100000	5040	9592	0.52543
200000	9485	17984	0.52741
300000	13789	25997	0.53040
400000	17985	33860	0.53115
500000	22142	41538	0.53305
600000	26226	49098	0.53415
700000	30274	56543	0.53541
800000	34260	63951	0.53572
900000	38236	71274	0.53646
1000000	42154	78498	0.53700
2000000	80559	148933	0.54090
3000000	117742	216816	0.54305
4000000	154219	283146	0.54466
5000000	190177	348513	0.54568
6000000	225817	412849	0.54697
7000000	260974	476648	0.54751
8000000	295901	539777	0.54819
9000000	330648	602489	0.54880
10000000	365128	664579	0.54941

The following graph shows that the ratio $L_2(n)/L_1(n)$ seems to have a limiting value below 0.56 when n tends to ∞. In any case, asymptotically for any $k-$value it is plausible that $L_k(n) = c_k L_1(n)$ where $1 > c_2 > c_3 > \cdots > c_k > \ldots$ are unknown positive constants.



It is simple to see that each $A(k)-$sequence is infinite since if it were finite with a last term $a_{k,n}$ then studying of the number $a_{k,1}a_{k,2}\ldots a_{k,n}+k$ leads immediately to a contradiction.

Update 24 Aug 2005:

I have computed the frequencies $n_k = L_k(2\times 10^9)$ for $k = 1, 2, \ldots, 10$ after making a simple sieve program in Survo. The following summary tells then lower limits of constants c_k as n_k/n_1. Sloane's Encyclopedia now includes these sequences.

k	n_k	n_k/n_1	Sloane
1	98222287	1	A000040
2	56472931	0.5750	A109022
3	34281318	0.3490	A109328
4	28159808	0.2867	A109329
5	19810400	0.2017	A109330
6	18346769	0.1868	A109331
7	14786395	0.1505	A109332
8	13281120	0.1352	A109333
9	11429212	0.1164	A109334
10	10921870	0.1112	A109335

Update 30 Aug 2005:

The maximum gap between $A(2)-$numbers below 2×10^9 is 513 (from 1743756419 to 1743756932) while that for $A(1)-$numbers (primes) is 292.

The number of $A(2)-$pairs with the minimal gap 3 is 1343185 and their proportion of all $A(2)-$numbers in the same range is 0.0238 while the corresponding ratio for twin primes is 0.0650.

Ratios $r_{k,n} = n_k/n_1$ are very crude lower limits for the c_k numbers. I have tried to study their asymptotic behaviour when $k = 2$.

A model of the form

M1: $r_{2,n} = a - b/\ln n, \quad a \equiv c_2$

seems to work better than others having the same simplicity. By numerical experiments done by the ESTIMATE program of the Survo system and by aiming at a good predicting property I have generalized this model stepwise to forms

M2: $r_{2,n} = a - b/\ln\ln n$
M3: $r_{2,n} = a - b\ln\ln n/\ln n$
M4: $r_{2,n} = a - b\ln\ln n\ln\ln\ln n/\ln n$

The parameters a, b and their standard errors s_a, s_b have been estimated from values $r_{2,n},\ n = 10^6(10^6)10^9$. Then the predicted values of $r_{2,n}$ have been computed for $n = 10^9 + 10^6(10^6)2 \times 10^9$.

The following summary includes the mean and the standard deviation of the predicted values as well as the minimum and maximum prediction errors. The last model (M4) is clearly the best one but it should be observed that all these model underestimate the true values.

					Prediction	error		
	a	s_a	b	s_b	mean	std.dev.	min	max
M1	0.6548	0.00025	1.71026	0.0049	0.000747	0.000170	0.000415	0.001031
M2	0.8306	0.00063	0.78594	0.0014	0.000606	0.000137	0.000334	0.000838
M3	0.7017	0.00028	0.89078	0.0012	0.000520	0.000119	0.000283	0.000722
M4	0.8326	0.00013	1.60882	0.0008	0.000061	0.000014	0.000021	0.000093

If parameters a, b are estimated from the last 1000 observations (the predicted cases above) the growth of the a is 2.7 per cent according to M1 but only 0.4 per cent by M4. M4 indicates that $a \equiv c_2$ would be about 0.85 but the ratio $r_{2,n}$ would grow very slowly. The model M4 gives, for example,

$r_{2,10^{12}} = 0.6012$
$r_{2,10^{15}} = 0.6246$
$r_{2,10^{100}} = 0.7708$
$r_{2,10^{200}} = 0.7966$
$r_{2,10^{300}} = 0.8070$

The models and results given by them are pure guesses without any theoretical framework. Therefore one must be very cautious when making any conclusions.

Update 3 Sep 2005: Sequences with mutual residues exactly k

We define another integer sequence $B(k)$

$$b_{k,1},\ b_{k,2},\ \ldots,\ b_{k,n}, \ldots$$

with mutual residues (exactly) k as

$$b_{k,1} = k + 1, \quad b_{k,n} = \min\{m \mid m > b_{k,n-1},\ \mathrm{mod}(m, b_{k,i}) = k,\ i = 1, 2, \ldots, n-1\}.$$

This more stringent requirement leads to well-known sequences. The general term in all cases $k = 1, 2, \ldots$ is

$$b_{k,n} = b_{k,1}b_{k,2}\ldots b_{k,n-1} + k.$$

To prove this, it is necessary and sufficient to show that terms in the sequence are coprimes. This is done by induction and by using the first step of the Euclidean Algorithm repeatedly. It is clear that for $b_{k,1} = k + 1$ and $b_{k,2} = 2k + 1$ we have $\gcd(b_{k,1}, k) = 1$, $\gcd(b_{k,2}, k) = 1$, and $\gcd(b_{k,2}, b_{k,1}) = 1$. By assuming that $\gcd(b_{k,i}, k) = 1$ and $\gcd(b_{k,i}, b_{k,j}) = 1$ for $i = 1, 2, \ldots, n-1$ and all $j < i$ it will be shown that $b_{k,n} = b_{k,1}b_{k,2}\ldots b_{k,n-1} + k$ has the same properties.

Since $\gcd(b_{k,i}, k) = 1,\ i = 1, 2, \ldots, n-1$ also $\gcd(b_{k,1}b_{k,2}\ldots b_{k,n-1}, k) = 1$ and $\gcd(b_{k,1}b_{k,2}\ldots b_{k,n-1}+k, k) = 1$. Thus $\gcd(b_{k,n}, k) = 1$. Similarly $\gcd(b_{k,n}, b_{k,i}) = \gcd(b_{k,1}b_{k,2}\ldots b_{k,n-1} + k, b_{k,i}) = \gcd(b_{k,i}, k) = 1$ for $i = 1, 2, \ldots, n-1$.

The sequences $B(k)$ obtained for values $k = 1, 2, \ldots, 10$ are

$k = 1$: Sylvester's sequence A000058
2, 3, 7, 43, 1807, 3263443, 10650056950807, . . .

$k = 2$: Fermat sequence A000215
3, 5, 17, 257, 65537, 4294967297, 18446744073709551617, . . .

$k = 3$: A000289
4, 7, 31, 871, 756031, 571580604871, 326704387862983487112031, . . .

$k = 4$: A000324
5, 9, 49, 2209, 4870849, 23725150497409, 562882766124611619513723649, . . .

$k = 5$: A001543
6, 11, 71, 4691, 21982031, 483209576974811, 233491495280173380882643611671, . . .

$k = 6$: A001544
7, 13, 97, 8833, 77968897, 6079148431583233, 36956045653220845240164417232897,
. . .

$k = 7$: A067686
8, 15, 127, 15247, 232364287, 53993160246468367, 2915261353400811631533974206368127,
. . .

$k = 8$: A110360 (4 Sep 2005)
9, 17, 161, 24641, 606981761, 368426853330807041, 1357383462555240000293762417728719361,
18424898644107427010977107148874723523180059431182608785043639266493441, . . .

$k = 9$: A110368 (4 Sep 2005)
10, 19, 199, 37819, 1156948199, 1654362331095061619,
2736914722546286269314723509551346599,
7490702198490615150126275937342974843521061335534838392096463268266747419,
. . .

$k = 10$: A110383 (4 Sep 2005)
11, 21, 241, 55681, 3099816961, 9608865160705105921, 92330289676612360941221747472778199041,
8524882391767151111154918892947398067446166736305624023874497267723631329281,
...

An equivalent expression for $b_{k,n}$ is ([1],[2])

$$b_{k,n} = b_{k,n-1}^2 - kb_{k,n-1} + k.$$

Update 10 Sep 2005: Sequences with mutual residues $-k$

Another family related to sequences $A(k)$ and more closely to $B(k)$ is the following one. All the terms in the sequence $C(k)$ should have mutual residues $-k$ and thus it is defined as

$$c_{k,1} = k+1, \quad c_{k,n} = \min\{m \mid m > c_{k,n-1},\ \mathrm{mod}(m, b_{k,i}) = -k,\ i = 1, 2, \ldots, n-1\}.$$

Possible values for k are $1, 2, \ldots$. The two first terms of the $C(k)-$sequence are $c_{k,1} = k+1, c_{k,2} = 2k+1$ by definition. The general term for $n = 3, 4, \ldots$ is

$$c_{k,n} = c_{k,1}c_{k,2}\ldots c_{k,n-1} - k.$$

which follows from the fact that all the terms are coprimes. The proof is similar to that for the $B(k)-$sequences.

It is also true that

$$c_{k,n} = c_{k,n-1}^2 + kc_{k,n-1} - k, \quad n = 4, 5, \ldots$$

since

$$\begin{aligned} c_{k,n} &= c_{k,1}c_{k,2}\ldots c_{k,n-1} - k \\ &= c_{k,n-1}(c_{k,1}c_{k,2}\ldots c_{k,n-2} - k + k) - k \\ &= c_{k,n-1}(c_{k,n-1} + k) - k \\ &= c_{k,n-1}^2 + kc_{k,n-1} - k. \end{aligned}$$

The sequences $C(k)$ obtained for values $k = 1, 2, \ldots, 10$ are

$k = 1$: A110389 (11 Sep 2005)
2, 3, 5, 29, 869, 756029, 571580604869, 326704387862983487112029,
106735757048926752040856495274871386126283608869, ...
This is identical to A005267 in [3] except that the two first terms are in reverse order. A005267 starts by $3, 2, 5, 29, 869, \ldots$.

$k = 2$: A110407 (11 Sep 2005)
3, 5, 13, 193, 37633, 1416317953, 2005956546822746113, 4023861667741036022825635656102100993,
16191462721115671781777559070120513664958590125499158514329308740975788033,
...

$k = 3$: A110413 (11 Sep 2005)
4, 7, 25, 697, 487897, 238044946297, 56665396458255748851097,

3210967155771303165846414430093064202724656697, . . .

$k = 4$: A110421 (11 Sep 2005)
5, 9, 41, 1841, 3396641, 11537183669441, 133106607022462246291930241,
17717368833032195779538884761310335951434822778039041, . . .

$k = 5$: A110445 (11 Sep 2005)
6, 11, 61, 4021, 16188541, 262068940651381,
68680129654138367181280464061,
4716960209309256311616420732713790878862755260077914932021, . . .

$k = 6$: A110455 (11 Sep 2005)
7, 13, 85, 7729, 59783809, 3574104177251329,
12774220669845420831090695774209,
163180713721905992070758583926701857930269220543803914084220929, . . .

$k = 7$: A110459 (11 Sep 2005)
8, 15, 113, 13553, 183778673, 33774601936091633,
1140723735941444920716624925248113,
1301250641740207358399613061389386702544244320473228561469086797553, . . .

$k = 8$: A110462 (11 Sep 2005)
9, 17, 145, 22177, 491996737, 242060793154621057,
58593427582644242304230418504765697,
3433189755882575096320786725947643765954772037072168669374448906021377, . . .

$k = 9$: A110463 (11 Sep 2005)
10, 19, 181, 34381, 1182362581, 1397981283590244781,
1954351669268628414383088499809940981,
3819490447173074341052986454970501004004783894423555148483555928992711181,
. . .

$k = 10$: A110466 (11 Sep 2005)
11, 21, 221, 51041, 2605694081, 6789641669815375361,
46099234004493318683404288695479633921,
2125139375801033098865842355143570823564637490880685503089270581842970173441,
. . . .
The current version of this paper can be downloaded from
`http://www.survo.fi/papers/resseq.pdf`

References

[1] A. V. Aho and N. J. A. Sloane, Some doubly exponential sequences, Fib. Quart., 11 (1973),
[2] S. W. Golomb, On certain nonlinear recurring sequences, Amer. Math. Monthly 70 (1963), 403-405.
[3] N. J. A. Sloane, "The On-Line Encyclopedia of Integer Sequences." `http://www.research.att.com/~njas/sequences`
[4] `http://www.survo.fi/english`

Department of Mathematics and Statistics, University of Helsinki
E-mail address: `seppo.mustonen@helsinki.fi`