How often does

$2kp+1$ divide p^p-1?

Sangil Nahm
Purdue University Math PhD, 2011
West Lafayette, Indiana
(now with Samsung)

Sam Wagstaff
Purdue University CERIAS
West Lafayette, Indiana

February 24, 2020

Computational Number Theory

In 1962, R. W. Hamming wrote:

The purpose of computing is insight, not numbers.

The ideal piece of work in computational number theory:

Write a program.

Run it.

Look at the output and discover new theorems.

Sangil's thesis dealt with Bell numbers, which arise in combinatorics. Define $B(n)$ to be the number of ways a set of size n can be partitioned into the disjoint union of 1 or more (non-empty) subsets.

Example: The set $\{1, 2, 3\}$ of size 3 can be partitioned as

- $\{1\} \cup \{2\} \cup \{3\}$,
- $\{1, 2\} \cup \{3\}$,
- $\{1, 3\} \cup \{2\}$,
- $\{1\} \cup \{2, 3\}$, or
- $\{1, 2, 3\}$,

so $B(3) = 5$.

The first few Bell numbers for $n = 1, 2, \ldots$ are 1, 2, 5, 15, 52, 203, 877, 4140, 21147, $\ldots$. (By convention, $B(0) = 1$.)

They are named after E. T. Bell [1934], but were first studied by Ramanujan in his (unpublished) notebook 20 years earlier.

The thesis investigates the minimum period of the Bell numbers $B(n)$ reduced modulo a prime p.

Background

J. Touchard's congruence [1933]

$$B(n+p) \equiv B(n) + B(n+1) \bmod p,$$

valid for any prime p and for all $n \geq 0$, shows that any p consecutive values of $B(n) \bmod p$ determine the sequence modulo p after that point.

Example. The sequence $B(n) \bmod 3$ for $n \geq 0$ is
1, 1, 2, 2, 0, 1, 2, 1, 0, 0, 1, 0, 1,
1, 1, 2, 2, 0, ... , with period length 13.

It follows from this congruence that $B(n) \bmod p$ must be periodic with period $\leq p^p$.

In 1945, G. T. Williams proved that for each prime p, the period of the Bell numbers modulo p divides

$$N_p = (p^p - 1)/(p - 1).$$

In fact the minimum period equals N_p for every prime p for which this period is known.

Theorem 1. The minimum period of the sequence $\{B(n) \bmod p\}$ is N_p when p is a prime < 126 and also when $p = 137$, 149, 157, 163, 167 or 173.

This theorem is proved by showing that the period does not divide N_p/q for any prime divisor q of N_p.

If q divides N_p and $N = N_p/q$, then one can test whether the period of the Bell numbers modulo p divides N by checking whether $B(N+i) \equiv B(i) \bmod p$ for $0 \le i \le p-1$. The period divides N if and only if all p of these congruences hold.

A polynomial time algorithm for computing $B(n) \bmod p$ is known.

The theorem for p can be proved (or disproved) this way if we know the factorization of N_p.

It is conjectured that the minimum period of the Bell numbers modulo p equals N_p for every prime p.

The conjecture is known to be true for all primes < 126 and for a few larger primes.

We give a heuristic argument for the probability that the conjecture holds for a prime p.

The most difficult piece of this heuristic argument is determining the probability that a prime q divides N_p. This probability is studied in this talk.

How often does $2kp+1$ divide N_p as p varies?

It is well known (Euler, 1755) that when p is prime every prime factor of N_p has the form $2kp+1$.

For each $1 \leq k \leq 50$ and for all odd primes $p < 100000$, we computed the fraction of the primes $q = 2kp+1$ that divide N_p.

For example, when $k = 5$ there are 1352 primes $p < 100000$ for which $q = 2kp+1$ is also prime, and 129 of these q divide N_p, so the fraction is $129/1352 = 0.095$.

This fraction is called “Prob” in the table because it approximates the probability that q divides N_p, given that p and $q = 2kp+1$ are prime, for fixed k.

Probability that $q = (2kp + 1) \mid N_p$

k	Prob
1	0.503
2	1.000
3	0.171
4	0.247
5	0.095
6	0.173
7	0.076
8	0.496
9	0.047
10	0.096
11	0.042
12	0.082
13	0.051
14	0.068
15	0.033
16	0.064
17	0.032
18	0.111
19	0.021
20	0.050

Probability that $q = (2kp+1) \mid N_p$

k	$1/k$	Prob
1	1.000	0.503
2	0.500	1.000
3	0.333	0.171
4	0.250	0.247
5	0.200	0.095
6	0.167	0.173
7	0.143	0.076
8	0.125	0.496
9	0.111	0.047
10	0.100	0.096
11	0.091	0.042
12	0.083	0.082
13	0.077	0.051
14	0.071	0.068
15	0.067	0.033
16	0.063	0.064
17	0.059	0.032
18	0.056	0.111
19	0.053	0.021
20	0.050	0.050

Probability that $q = (2kp+1) \mid N_p$					
Odd k			Even k		
k	$1/(2k)$	Prob	k	$1/k$	Prob
1	0.500	0.503	2	0.500	1.000
3	0.167	0.171	4	0.250	0.247
5	0.100	0.095	6	0.167	0.173
7	0.071	0.076	8	0.125	0.496
9	0.056	0.047	10	0.100	0.096
11	0.045	0.042	12	0.083	0.082
13	0.038	0.051	14	0.071	0.068
15	0.033	0.033	16	0.063	0.064
17	0.029	0.032	18	0.056	0.111
19	0.026	0.021	20	0.050	0.050
21	0.024	0.016	22	0.045	0.054
23	0.022	0.021	24	0.042	0.042
25	0.020	0.021	26	0.038	0.052
27	0.019	0.021	28	0.036	0.036
29	0.017	0.022	30	0.033	0.031
31	0.016	0.019	32	0.031	0.055
49	0.010	0.014	50	0.020	0.043

Obervations about the table

1. Prob is approximately $1/(2k)$ when k is odd.

2. Usually Prob is approximately $1/k$ when k is even.

3. Some anomalies to 2. are that Prob is about $2/k$ when $k = 2$, 18, 32 and 50.

4. Also, Prob is about $4/k$ when $k = 8$.

5. The exceptional values of k in 3. and 4. have the form $2m^2$ for $1 \leq m \leq 5$. (These numbers also arise as the lengths of the rows in the periodic table of elements in chemistry.)

We will now explain these observations. Suppose k is a positive integer and that both p and $q = 2kp + 1$ are odd primes. Let g be a primitive root modulo q.

If $p \equiv 1 \bmod 4$ or k is even (so $q \equiv 1 \bmod 4$), then by the Law of Quadratic Reciprocity

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) = \left(\frac{2kp+1}{p}\right) = \left(\frac{1}{p}\right) = +1,$$

so p is a quadratic residue modulo q. In this case $g^{2s} \equiv p \bmod q$ for some s. Now by Euler's criterion for power residues, $(2kp+1) \mid (p^p - 1)$ if and only if p is a $(2k)$-ic residue of $2kp+1$, that is, if and only if $(2k) \mid (2s)$. It is natural to assume that $k \mid s$ with probability $1/k$ because k is fixed and s is a random integer.

If $p \equiv 3 \bmod 4$ and k is odd (so $q \equiv 3 \bmod 4$), then

$$\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right) = -\left(\frac{2kp+1}{p}\right) = -\left(\frac{1}{p}\right) = -1,$$

so p is a quadratic nonresidue modulo q. Now $g^{2s+1} \equiv p \bmod q$ for some s. Reasoning as before, $(2kp+1) \mid (p^p - 1)$ if and only if $(2k) \mid (2s+1)$, which is impossible. Therefore q does not divide N_p.

Thus, if we fix k and let p run over all primes, then the probability that $q = 2kp+1$ divides N_p is $1/k$ when k is even and $1/(2k)$ when k is odd because, when k is odd only those $p \equiv 1 \bmod 4$ (that is, half of the primes p) offer a chance for q to divide N_p.

In fact, when $k = 1$ and $p \equiv 1 \bmod 4$, q always divides N_p. This theorem must have been known long ago, but we could not find it in the literature.

Theorem 2. If p is odd and $q = 2p+1$ is prime, then q divides N_p if and only if $p \equiv 1 \bmod 4$.

Proof. We have just seen that q does not divide N_p when $p \equiv 3 \bmod 4$. If $p \equiv 1 \bmod 4$, then p is a quadratic residue modulo q, as was mentioned above, so $p^p = p^{(q-1)/2} \equiv +1 \bmod q$ by Euler's criterion. Finally, q is too large to divide $p-1$, so q divides N_p.

We now explain the anomalies, beginning with $k = 2$.

Theorem 3. If $q = 4p + 1$ is prime, then q divides N_p.

This result was an ancient problem posed and solved more than 100 years ago. Here is a modern proof.

Proof. Since $q \equiv 1 \bmod 4$, there exists an integer i with $i^2 \equiv -1 \bmod q$. Then

$$(1+i)^4 \equiv (2i)^2 \equiv -4 \equiv \frac{1}{p} \bmod q.$$

Hence

$$p^p \equiv \left(\frac{1}{p}\right)^{-p} \equiv (1+i)^{-4p} \equiv (1+i)^{1-q} \equiv 1 \bmod q$$

by Fermat's theorem. Thus, q divides $p^p - 1$. But $q = 4p + 1$ is too large to divide $p - 1$, so q divides N_p.

Theorem 4. Let p be an odd positive integer and m be a positive integer. If $q = 4m^2p + 1$ is prime, then q divides $p^{m^2p} - 1$.

Of course, Theorem 3 is the case $m = 1$ of Theorem 4.

Fermat's theorem says that q divides $p^{4m^2p} - 1$.

In the case $m = 2$, that is, $k = 8$, we can do even better.

Fermat's theorem says that q divides $p^{16p} - 1$.

Theorem 5 If $q = 16p + 1$ is prime, then q divides $p^{2p} - 1$.

Proof. As in the proof of the previous theorem, we have i with $i^2 \equiv -1 \bmod q$ and $(1+i)^4 \equiv -4 \bmod q$. Therefore, $(1+i)^8 \equiv 16 \equiv -1/p \bmod q$ and so

$$p^{2p} \equiv (1+i)^{-16p} \equiv (1+i)^{1-q} \equiv 1 \bmod q,$$

which proves the theorem.

Thus, a prime $q = 2kp + 1 = 16p + 1$ divides $(p^p - 1)(p^p + 1)$ when $k = 8$. Assuming that q has equal chance to divide either factor, the probability that q divides $p^p - 1$ is $1/2$.

So far, we have explained all the behavior seen in the table. Further experiments with $q = 2m^2p + 1$ lead us to the following result, which generalizes Theorems 4 and 5.

The theorem lets us remove arbitrarily large powers of 2 from the exponent in certain cases.

Theorem 6. [Nahm and Montgomery] Suppose p, m, t are positive integers, with t a power of 2 and $t > 1$. Let $k = (2m)^t/2$ and $q = 2kp + 1 = (2m)^t p + 1$. If q is prime, then t divides k and $p^{kp/t} \equiv 1 \bmod q$.

When $t = 2$, the theorem is just Theorem 4.

When $t = 4$, Theorem 6 says that if $q = (2m)^4 + 1 = 16m^4 + 1$ is prime, then q divides $p^{2m^4 p} - 1$. Theorem 5 is the case $m = 1$ of this statement.

When $t = 8$, Theorem 6 says that if $q = (2m)^8 + 1 = 256m^8 + 1$ is prime, then q divides $p^{16m^8 p} - 1$. The first case, $m = 1$, of this statement is for $k = 128$, which is beyond the end of the table.

We now apply Theorem 6. As above, let g be a primitive root modulo q and let $a = g^{(q-1)t/k} \bmod q$. Then a^j, $0 \le j < k/t$, are all the solutions to $x^{k/t} \equiv 1 \bmod q$. Let $b = p^p \bmod q$. By the theorem, $b^{k/t} \equiv 1 \bmod q$, so $b \equiv a^j \bmod q$ for some $0 \le j < k/t$. It is natural to assume that $j = 0$, that is, $q \mid N_p$, happens with probability $1/(k/t) = t/k$.

Summary

We have given heuristic arguments which conclude that, for fixed k, when p and $q = 2kp + 1$ are both prime, the probability that q divides N_p is $c(k)/k$, where $c(k)$ is defined as follows:

When k is an odd positive integer, $c(k) = 1/2$.

When k is an even positive integer, let t be the largest power of 2 for which there exists an integer m so that $2k = (2m)^t$. Then $c(k) = t$.

We have

$$c(k) = \begin{cases} 1/2 & \text{if } k \text{ is odd,} \\ 1 & \text{if } k \text{ is even and } k \neq 2m^2, \\ O(\log k) & \text{if } k = 2m^2 \text{ for some } m. \end{cases}$$

The average value of $c(k)$ is 3/4 because the numbers $2m^2$ are rare.

Is the conjecture about the Bell numbers' period true? Does it always equal N_p?

Applying the Bateman-Horn conjecture, the Prime Number Theorem, the Binomial Theorem and the divisibility results for N_p, one can show that the heuristic probability that the minimum period of the Bell numbers modulo p is N_p is

$$(1 - p^{-p})^{3(\log p)/2} \approx 1 - \frac{3 \log p}{2p^p},$$

exceedingly close to 1 when p is large.

Finally, we compute the expected number of primes $p > x$ for which the conjecture fails. When $x > 2$, this number is

$$\sum_{p>x} \frac{3 \log p}{2p^p} < \sum_{p>x} p^{1-x} \leq \int_x^\infty t^{1-x}\, dt = \frac{x^{2-x}}{x-2}.$$

By Theorem 1, the conjecture holds for all primes $p < 126$. Taking $x = 126$, the expected number of primes for which the conjecture fails is $< 126^{-124}/124 < 10^{-262}$. Thus, the heuristic argument predicts that the conjecture is almost certainly true.

Math. Comp. **79** (2010) pp. 1793-1800.