Primality Testing and Carmichael numbers

Andrew Granville

The problem of distinguishing prime numbers from composite numbers is one of the most fundamental and important in arithmetic. It has remained as a central question in our subject from ancient times to this day[1], and yet still fascinates and frustrates us all. From the very definition of primality, that an integer

$$n \text{ is prime if it has no divisor between 2 and } \sqrt{n},$$

one can evolve a simple test for primality: Just check whether any integer d between 2 and $\sqrt{n}$ actually divides n. This is an easily implemented test for, say, $n = 107$ or $n = 11035$, but how about for $n = 1234567890123456 77$? This requires over a billion test divides, and if one were to try to verify that a given 100 digit integer n is prime in this way it would take longer than the remaining lifespan of our universe, even on an impossibly fast computer!

One thus needs a more sophisticated approach to handle large numbers. Perhaps a different definition of prime numbers will furnish us with a quicker method ? One such definition follows from Wilson's Theorem (1770):

$$n \text{ is prime if and only if } n \text{ divides } (n-1)! + 1.$$

So to find out whether n is prime we multiply together all integers less than n, add 1, and see whether the resulting number is divisible by n. However this requires multiplying $n-3$ pairs of numbers together, as opposed to $\sqrt{n}$ test divides earlier, so takes even longer than our previous method.

The ancient Chinese made the startling discovery that

$$\text{If } n \text{ is prime then } n \text{ divides } 2^n - 2,$$

which implies that

(1) If n does not divide $2^n - 2$ then n is composite (that is, not prime).

So we now have a new, and quite different, criterion, which will tell us that certain numbers n are composite. However, if a number fails this criterion (that is, if n does divide $2^n - 2$), then it doesn't, a priori, tell us that n is prime; but let's check it out:

2 divides $2^2 - 2 = 2$	...
3 divides $2^3 - 2 = 6$	101 divides $2^{101} - 2$
4 doesn't divide $2^4 - 2 = 14$	103 divides $2^{103} - 2$
5 divides $2^5 - 2 = 30$	105 doesn't divide $2^{105} - 2$
6 doesn't divide $2^6 - 2 = 62$	107 divides $2^{107} - 2$
7 divides $2^7 - 2 = 126$	109 divides $2^{109} - 2$
8 doesn't divide $2^8 - 2 = 254$	111 doesn't divide $2^{111} - 2$
9 doesn't divide $2^9 - 2 = 510$	etc.

[1] from Article 329 of Gauss's *Disquisitiones Arithmeticae* (1801)

In all of these examples we observe that n is prime exactly when n divides $2^n - 2$, and is composite otherwise. According to E. T. Bell, the ancient Chinese thought that this is always true [2], as did Leibniz many centuries later. However the (smallest such) example, $n = 341$, refutes this belief since $341 = 11 \times 31$ is composite, yet 341 divides $2^{341} - 2$.

Further computation shows that such composite n seem to be rare and so we define composite number n to be a **base 2 pseudoprime** if n divides $2^n - 2$. To exhibit quite how rare these are, note that up to 10^{10} there are around 450 million primes, but only about fifteen thousand such base 2 pseudoprimes, while up to $2 \cdot 5 \times 10^{10}$ there are over a billion primes, and yet fewer than 22 thousand base 2 pseudoprimes. So, if you were to choose a random number $n < 2 \cdot 5 \times 10^{10}$ for which n divides $2^n - 2$ then there would be a less than one-in-fifty-thousand chance that your number would be composite.

Testing whether $2^{n-1} \equiv 1 \bmod n$ is easily implemented on a computer, as follows:

(i) Write $n-1$ in base 2, say $n - 1 = 2^{a_k} + 2^{a_{k-1}} + \ldots + 2^{a_1}$ where $a_k > a_{k-1} > \ldots > a_1$

(ii) Compute $r_j \equiv 2^{2^j} \bmod n$ for $0 \le j \le a_k$, by taking $r_0 = 2$ and $r_{j+1} \equiv r_j^2 \bmod n$ for each $j \ge 0$

(iii) Finally, since $2^{n-1} = 2^{2^{a_k}} \cdot 2^{2^{a_{k-1}}} \ldots 2^{2^{a_1}}$, we have $2^{n-1} \equiv r_{a_k} r_{a_{k-1}} \ldots r_{a_1} \pmod n$.

This algorithm requires no more than $20 \log^3 n$ operations so that, for a 40 digit number n, this 'pseudoprime test' takes a few million operations (a few seconds on a PC) whereas test division takes more than a billion billion operations (over a thousand years on a PC). It has been suggested that one might obtain a practical primality test by writing down a list of all base 2 pseudoprimes, and then, if n divides $2^n - 2$ but is not on the list, one knows that n is a prime. Since there's less than 22 thousand base 2 pseudoprimes up to $2 \cdot 5 \times 10^{10}$, this method works well in this range, and will continue to work well as long as the base 2 pseudoprimes remain so scarce. However, this won't always be so since Malo proved, in 1903, that there are infinitely many odd composite base 2 pseudoprimes, by showing that if $n = ab$ (with $a, b > 1$) is such a number, then so is $n' = 2^n - 1$ [3]. This is proved by observing that, since a divides n which divides $n' - 1$, thus $x^a - 1$ divides $x^n - 1$, which divides $x(x^{n'-1} - 1) = x^{n'} - x$, and so, in particular with $x = 2$, we get that $2^a - 1$ divides $2^n - 1 = n'$ which divides $2^{n'} - 2$.

Our hope of obtaining a complete list of base 2 pseudoprimes is thus doomed, but we might still find all base 2 pseudoprimes up to some large number x. However, in 1982, Pomerance showed that there are more than $e^{(\log x)^c}$ base 2 pseudoprimes $\le x$, for some

[2] However it is now believed that Bell had no evidence of this, but was embellishing a good story: Just as standards of mathematical rigor have greatly improved over the last hundred years, so too the standards of rigor of mathematical history.

[3] and then we get the sequence n, $2^n - 1$, $2^{2^n - 1} - 1$, $2^{2^{2^n - 1} - 1} - 1, \ldots$ of base 2 pseudoprimes by iterating this observation.

constant c, $0 < c < 1$, once x is sufficiently large[4]. This is quite a fast growing function of x and shows that our hoped for, easy and quick primality test won't be practical for large values of x. So what else can we do ?

On October 18th, 1640 Fermat wrote, in a letter to his confidante Frenicle, that the fact that n divides $2^n - 2$ whenever n is prime is not an isolated phenomenon. Indeed that, if n is prime then

(2) $$n \text{ divides } a^n - a \text{ for all integers } a;$$

which implies that

If n doesn't divide $a^n - a$ for some integer a then n is composite.

So instead of considering pseudoprimes to base 2, we can consider pseudoprimes to any base a: it turns out that such pseudoprimes are rare, though some do exist. However, since base 2 pseudoprimes are rare, and base 3 pseudoprimes are also rare, one would guess that numbers that are both base 2 and base 3 pseudoprimes must be extremely rare; perhaps none exist at all ? Unfortunately some do exist, such as 2701, which divides both $2^{2701} - 2$ and $3^{2701} - 3$, yet $2701 = 37 \times 73$ is composite. Numbers that are pseudoprimes to bases 2, 3 and 5 simultaneously should be even rarer, but again do exist, for instance $n = 181 \times 361$; and, indeed, there are examples for any finite set of bases. So maybe we should ask whether there are any composite numbers n which are pseudoprime for every base a ? That is, for which (2) holds. Such a number n would have to have certain extraordinary properties:

(i) n must be squarefree, else if p^2 divides n then $p^2 | n | p^n - p$ which is false.

(ii) If prime p divides n then $p - 1$ must divide $n - 1$, for if a is a primitive root mod p then a has order $p - 1$ mod p, but $a^{n-1} \equiv 1 \pmod{p}$ by (2).

In 1899 Korselt[5] observed that these two conditions imply that (2) holds (which the reader may verify – hint: use the Chinese Remainder Theorem). We thus state

Korselt's criterion: n divides $a^n - a$ for all integers a if and only if n is squarefree and $p - 1$ divides $n - 1$ for all primes p dividing n.

So now, to determine whether (2) holds for n, we need only verify a few simple properties of its prime factors. Korselt did not exhibit an example of such an integer n, and he might have thought that no such n exist. However such n do exist, as was discovered by Carmichael in 1910, the smallest being $561 = 3 \times 11 \times 17$. These numbers are now known as *Carmichael numbers*, but surely would have been known as Korselt numbers had he just done a few computations !

---

[4] for those readers not accustomed to such 'estimates', we note that, $e^{(\log x)^c}$ is larger than any given power of $\log x$, and smaller than any given (positive) power of x, for sufficiently large x.

[5] responding to a 'Problème Chinois' from *L'Intermédiaire des Mathématiciens*, a turn-of-the-century French journal, similar to today's *The American Mathematical Monthly*

The first few Carmichael numbers are

$$\begin{aligned} 561 &= 3 \times 11 \times 17 \\ 1105 &= 5 \times 13 \times 17 \\ 1729^* &= 7 \times 13 \times 19 \\ 2465 &= 5 \times 17 \times 29 \\ 2821 &= 7 \times 13 \times 31 \end{aligned}$$

Notice how they all have three prime factors. To obtain one with four prime factors we must go out to

$$41041 = 7 \times 11 \times 13 \times 41$$

and for five prime factors to

$$825265 = 5 \times 7 \times 17 \times 19 \times 73.$$

Carmichael computed fifteen such numbers in his 1912 paper and stated that '*this list might be indefinitely extended*'. However it soon became apparent that it was going to be difficult to prove that his list could be so lengthened, and this statement has since been considered an open problem[6].

Korselt's criterion may be re-written as follows:

$n = p_1 p_2 \ldots p_k$ is a Carmichael number if and only if the p_i's are distinct and
$L = LCM[p_1 - 1, p_2 - 1, \ldots, p_k - 1]$ divides $n - 1$.

So, to verify that those numbers listed above are indeed Carmichael numbers, we only need check that $L = 80 = LCM[2, 10, 16]$ divides 560, that $L = 48$ divides 1104, that $L = 36$ divides 1728, that $L = 112$ divides 2464, that $L = 60$ divides 2820, that $L = 120$ divides 41040, and finally that $L = 144$ divides 825264. Notice that L is extremely small compared to $n - 1$ in each example, which gives us a hint as to how to find more Carmichael numbers: Let's try to find a set of primes where these primes minus one have

---

* 1729 is best-known from the story of when Hardy visited Ramanujan in hospital, and pronounced his taxicab number, 1729, to be a dull number. Ramanujan refuted this by noting that it is the smallest number which is the sum of two cubes in two different ways. However Ramanujan didn't say that 1729 is also interesting as being the third smallest Carmichael number! Carl Pomerance further observes that the second smallest Carmichael number, 1105, is the sum of two squares in more ways than any preceeding number. We leave it to the reader to come up with the analogous remark for 561, the smallest Carmichael number!

[6] see Alford's forthcoming paper *Chasing Carmichael numbers* for a revealing discussion of Carmichael's paper.

a surprisingly small common multiple. For example, since the prime divisors of 1729 are $p = 6 + 1,\ q = 12 + 1,\ r = 18 + 1$ giving $L = 36$, we can generalize this to

$$p = 6k + 1, \quad q = 12k + 1, \quad r = 18k + 1, \tag{3}$$

for integer $k \geq 1$, giving $L = 36k$. Since $pqr - 1 = 36k(36k^2 + 11k + 1)$, Korselt's criterion tells us that pqr is a Carmichael number provided each of $p,\ q$ and r are prime. It is easy to find many values of k for which the three numbers in (3) are simultaneously prime, but can we prove that there are infinitely many such k ? This is considered an outstandingly difficult open problem in analytic number theory, and although experts are certain that infinitely many such k do exist, there have been no plausible ideas as to how to prove such a result.

One can obtain other sequences in which one expects infinitely many prime triplets or quadruplets or quintuplets, which would give rise to infinitely many Carmichael numbers, for instance

$$(12k + 5)(36k + 13)(48k + 17), \qquad (6k + 7)(12k + 13)(18k + 19),$$
$$(28k + 5)(112k + 17)(196k + 29), \qquad (30k + 7)(60k + 13)(150k + 31),$$
$$(180k + 7)(300k + 11)(360k + 13)(1200k + 41);$$

but it seems unlikely that this approach will lead to a proof that there are infinitely many Carmichael numbers in the foreseeable future.

Let $C(x)$ be the number of Carmichael numbers up to x. The following table gives the number of Carmichael numbers up to various values of x:

x	$C(x)$	Year	Discoverer(s)
10^3	1	1910	Carmichael
10^4	7	1912	Carmichael
10^5	16		
10^6	43		
10^7	105		
10^8	255	1938	Poulet
10^9	646	1975	Swift
10^{10}	1547		
2.5×10^{10}	2163	1980	Pomerance, Selfridge, Wagstaff
10^{11}	3605		
10^{12}	8241	1990	Jaeschke
10^{13}	19279		
10^{14}	44706		
10^{15}	105212	1992	Pinch

This data suggests that there must indeed be infinitely many Carmichael numbers, even though they remain fairly scarce all the way up to 10^{15}. In 1949 Paul Erdős showed quite

how scarce Carmichael numbers are, by proving that the sum of their reciprocals converge [7]; it has since been proved that

$$(4)^{\dagger} \qquad C(x) \le x^{1-\{1+o(1)\}\log\log\log x/\log\log x}$$

In 1956 Erdős took a radically different approach to constructing Carmichael numbers: Earlier we noted that $L = \text{LCM}[p_1 - 1, \dots, p_k - 1]$ is much smaller than $n-1$ for most Carmichael numbers $n = p_1 \dots p_k$. However, for a typical set of primes, $\{p_1, \dots, p_k\}$, there is no particular reason to expect this to happen, indeed we'd expect L to be just a bit smaller than $n-1$. So to construct Carmichael numbers we must find some way of forcing L to be small. In our constructions above (like (3)), we selected our primes p to have certain special forms: this guaranteed that the $p-1$ had large common divisors, forcing L to be small compared to $n-1$. Erdős approached this problem from the other direction. Instead of choosing primes in special ways so as to force L to be small, he chose L so that there are many primes p for which $p-1$ divides L. Once this is done, one need only find a subset of these primes, say $p_1, p_2, \dots, p_k$, for which $n = p_1 p_2 \dots p_k \equiv 1 \pmod{L}$, to obtain the Carmichael number n — one sees that n is a Carmichael number, by using Korselt's criterion, since n is squarefree, and each $p_i - 1$ divides L, which divides $n-1$. Let's review

Erdős's construction of Carmichael numbers

(i) Select integer L;

(ii) Determine primes p for which $p-1$ divides L, but p does not divide L;

(iii) Find a subset of the primes obtained in (ii) whose product is $\equiv 1 \pmod{L}$. This product is a Carmichael number.

As an example, let's try (i) $L = 120$. The primes p which do not divide 120, but for which $p-1$ does, are (ii) 7, 11, 13, 31, 41, 61. Checking through all subsets of these primes we find that (iii) $41041 = 7 \times 11 \times 13 \times 41 \equiv 1 \bmod 120$, and $172081 = 7 \times 13 \times 31 \times 61 \equiv 1 \bmod 120$, and $852841 = 11 \times 31 \times 41 \times 61 \equiv 1 \bmod 120$, so that 41041, 172081 and 852841 are all Carmichael numbers.

With bigger, highly composite, values of L, we expect to find many more Carmichael numbers. Indeed if we obtain r different primes in step (ii) above, then there are $2^r - 1$ distinct products of non-trivial subsets of these primes. It seems plausible that roughly $1/L$ of these products are $\equiv 1 \pmod{L}$, and so we would have approximately $2^r/L$

[7] unlike the primes, whose sum of reciprocals diverge.

†For those not accustomed to such estimates, this is larger than $x^{1-\epsilon}$ for any fixed $\epsilon > 0$, but smaller than any given positive constant times x, once x is sufficiently large.

Carmichael numbers so formed. It can be shown that if L is the product of all the primes up to some sufficiently large point, then we can obtain more than $2\log^3 L$ primes in (ii), and so we'd expect more than $L^{\log^2 L}$ such Carmichael numbers. Erdős gave a similarly reasoned argument to justify his conjecture that for any fixed $\epsilon > 0$, there are more than $x^{1-\epsilon}$ Carmichael numbers up to x, once x is large enough [8].

However we see from our table above that the Carmichael numbers remain scarce all the way up to 10^{15}, which is surprising if Erdős's conjecture is to be believed. Indeed Dan Shanks, in his book *Solved and Unsolved problems in number theory*, challenged those who believe Erdős's conjecture to produce a value of x for which there are more than $x^{1/2}$ Carmichael numbers up to x. (Note that up to $x = 10^{15}$, there are only a few more than $x^{1/3}$ Carmichael numbers.)

It is important to note that Erdős's construction is impractical, both theoretically and computationally, if one doesn't know how to find products, of the primes produced in (ii), which are $\equiv 1 \pmod{L}$, as required for (iii). At the beginning of this year, there were fewer than ten thousand Carmichael numbers known, and it seemed to be a very difficult task to find many more. Then, suddenly on January 21st, 'Red' Alford announced that he had proven the existence of at least 2^{128} Carmichael numbers ! Unlike previous computations, which had sought all the Carmichael numbers up to some pre-assigned limit, or had found many in certain sequences (such as in that given by (3)), Alford modified Erdős's construction so as to make it computationally practical: As we've already discussed, it is easy (computationally) to implement steps (i) and (ii) above, but how can we find subsets of the primes in (ii) whose product is $\equiv 1 \bmod L$? Here's Alford's idea:

(iiia) Find a subset P of the primes in (ii), such that for every a, $1 \le a \le L$ with $\gcd(a, L) = 1$, there is a subset $p_1, \ldots, p_k$ of P for which $p_1 p_2 \ldots p_k \equiv a \pmod{L}$;

(iiib) Let Q be the primes found in (ii), excluding those belonging to P. For any subset $q_1, \ldots, q_r$ of these primes, let a be that integer, $1 \le a \le L$, which is $\equiv (q_1 q_2 \ldots q_r)^{-1} \pmod{L}$. From (iiia) we know that there is a subset $p_1, \ldots, p_k$ of P for which $p_1 \ldots p_k \equiv a \equiv (q_1 \ldots q_r)^{-1} \pmod{L}$, and so $p_1 \ldots p_k q_1 \ldots q_r \equiv 1 \pmod{L}$. Therefore, by Erdős's construction, $p_1 \ldots p_k q_1 \ldots q_r$ is a Carmichael number.

Thus, for each different non-trivial subset of Q we've constructed a different Carmichael number, providing a total of at least $2^{|Q|} - 1$ Carmichael numbers. This method is very practical, since we don't need to explicitly write down the Carmichael numbers constructed in (iiib) to be guaranteed of their existence; all we need know is that there is some product

[8] and, taking his argument to its limit, one expects $C(x)$ to be approximately the size of the function in (4)

of the primes in P in the congruence class $(q_1 q_2 \dots q_r)^{-1} \pmod{L}$ corresponding to each subset $q_1, \dots, q_r$ of Q.

It remains to find a suitable set P in (iiia). To do this, suppose that the primes found in (ii) were $p_1 < p_2 < \dots < p_m$, and define R_j to be the set of products $\pmod{L}$ of the subsets of $p_1, p_2, \dots, p_j$. We easily obtain R_{j+1} from R_j by observing that $R_{j+1} = R_j \cup \{rp_{j+1} \pmod{L} : r \in R_j\}$. Once we find j for which R_j is the set of all residue classes $a \pmod{L}$ with $1 \le a \le L$ and $(a, L) = 1$, then we can take $P = R_j$ and we're done.

Alford worked with the example (i) $L = 2^6 \cdot 3^3 \cdot 5^2 \cdot 7^2 \cdot 11$, and found that there are (ii) 155 primes $p \ge 13$ such that $p - 1$ divides L. By computing $R_1, R_2, \dots$ as above he got (iiia) $P = R_{27}$, that is that every residue class $a \pmod{L}$ with $(a, L) = 1$ is given by the product of some subset of the smallest 27 primes found in (ii). Thus if Q is the set of the largest $128 (= 155 - 27)$ primes found in (ii) then, as described above, each subset of Q corresponds to a Carmichael number, and we've proved the existence of at least $2^{128} - 1$ Carmichael numbers.

So, in an afternoon's work, Alford increased the number of Carmichael numbers known from fewer than 2^{14}, to more than 2^{128}. Certain faculty members, here at the University of Georgia, taunted the number theory group that there cannot be interesting finite sets which contain more than 2^{128} elements, and that surely Alford's idea should provide sufficient impetus to finally prove that there are infinitely many Carmichael numbers. And indeed it did. The theorem that we eventually proved is

Theorem. *(Alford, Granville, Pomerance – 1992): There are more than $x^{2/7}$ Carmichael numbers up to x, once x is sufficiently large.*

To make Erdős's construction theoretically practical, one evidently needs a result which guarantees that, given enough primes satisfying (ii), there is some subset whose product is $\equiv 1 \pmod{L}$. A theorem of van Emde Boas and Kruyswijk implies that if $m > 2$ is the largest order of an element of the multiplicative group modulo L, then such a subset exists provided there are more than $m \log L$ primes satisfying (ii). A theorem of Prachar guarantees the existence of integers L for which there are more than $L^{c/\log\log L}$ primes p satisfying (ii); however this quantity is usually a lot smaller than $m \log L$. To avoid this difficulty one wishes to select L so that m is very small, but Prachar's construction doesn't allow this. So instead we showed the existence of integers L of the form $L'k$ with $(L', k) = 1$, where the maximal order m' of an element modulo L' is extremely small, and there are more than $m' \log L$ primes p satisfying (ii), each with the additional property that $p \equiv 1 \pmod{k}$. The result of van Emde Boas and Kruyswijk then guarantees the existence of a subset of these primes whose product is $1 \pmod{L'}$ and, since any such product is $1 \pmod{k}$ (as each such prime is $1 \pmod{k}$), thus this product is $1 \pmod{L}$, and so a Carmichael number, from Erdős's construction.

Filling in the details of this outline involves some deep tools from analytic number theory, as well as combinatorial techniques involving groups and sets. This will all be described in detail in a forthcoming journal article.

One ingredient needed for the proof is a lower bound for the number of primes in certain arithmetic progressions: As is well known, there are asymptotically $x/\log x$ primes up to x, and we expect these to be more-or-less equally distributed amongst the arithmetic progressions $a \pmod d$ with $(a,d)=1$, provided d is a little smaller than x. Currently it is only known how to prove such a result if d is considerably smaller than x, in fact smaller than a fixed power of $\log x$. However, for our purposes, we proved

> *Fix $\epsilon > 0$. If x is sufficiently large then for all, but a few*[9], *integers $d \le x^{5/12-\epsilon}$ there are more than $x/2d\log x$ primes $\le x$ in the arithmetic progression $1 \pmod d$.*

It is widely believed that such a result holds for any $d \le x^{1-\epsilon}$. If true this implies Erdős's conjecture, for we also proved

Theorem. *(Alford, Granville, Pomerance – 1992): Fix $\epsilon > 0$. Assume that, for sufficiently large x, the arithmetic progression $1 \pmod d$ contains more than $x/2d\log x$ primes up to x provided $d \le x^{1-\epsilon}$. Then there are more than $x^{1-2\epsilon}$ Carmichael numbers up to x, once x is sufficiently large.*

This Theorem seems to guarantee that Erdős's conjecture is correct. So, in answer to Shanks's challenge to find an x for which $C(x) > x^{1/2}$, one can extrapolate our tabulated values of $\log C(x)/\log x$ to guess that one needs x to be around 10^{60} — it wouldn't be feasible to write down all the Carmichael numbers up to this point !!

So what does all this tell us about primality tests ? Although there are various methods known that will verify that a given number is prime in a 'small' number of steps (thanks to Miller, Goldwasser and Kilian, Adleman and Huang, and others), they all consist of checking a large number of conditions (polynomial in the number of digits of n). It would be more elegant if one only needed to check a finite number of such conditions, but it now seems unlikely that any such method proposed thus far will work.

In particular, there are various widely-used software packages that assert that a given integer is prime if it is a 'strong pseudoprime' for some given finite set of bases. However we can prove that, for any given finite set of bases, there are infinitely many Carmichael numbers that are 'strong pseudoprimes' to all the bases in that set. Such numbers would be falsely identified as prime by such a software packages, so reader, beware !

[9] A precise description of 'but a few' is: There exists an integer c (depending only on ϵ) such that there is a set B of no more than c integers, each $> \log x$, such that we must miss out all those d above that are divisible by an element of B.