Primes and Pseudoprimes

Primality tests and proofs

R.G.E. Pinch

Cheltenham
www.chalcedon.demon.co.uk

April 2010 / British Mathematical Colloquium

Number theory

Number theory ...

- ... is pure mathematics with applications
- ... is a theory which is also an experimental science
- ... reveals a deep order which generates randomness
- ... poses problems easy to state but hard to answer

The problem of distinguishing prime numbers from composite numbers and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic.
[...]
Further, the dignity of the science itself seems to require that every possible means be explored for the solution of a problem so elegant and so celebrated.

— C.F. Gauss
— Disquisitiones Arithmeticae §329

“Experimental evidence”

```
ifactor(3290922560713061);
```

(3290922560713061)

```
isprime(3290922560713061);
```

true

```
3290922560713061 mod 29363;
```

0

```
isprime(2432870015941);
```

true

```
ifactor(2432870015941);
```

(1213)(3637)(551461)

Trial division

The oldest test for primality is *trial division*: to determine the character of n, try possible factors t up to $\sqrt{n}$.

Disadvantages:

- **Slow**. In the worst case, n prime, takes $O\left(n^{1/2}\right)$ steps
- No way to certify when n is prime.

Properties of primes I

If p is prime then ...

- (Fermat) $b^{p-1} \equiv 1 \bmod p$ for all b coprime to p.
- (Euler–Jacobi) $b^{(p-1)/2} \equiv \pm 1 \equiv \left(\frac{b}{p}\right)_J \bmod p$ for all b coprime to p.
- (Miller–Rabin) $p = 2$ or $p - 1 = 2^r.s$ and the sequence

$$b^s, b^{2s}, \ldots, b^{2^r s} = b^{p-1} \bmod p$$

starts with 1 or contains $\ldots, -1, 1, \ldots$.

Properties of primes II

If p is prime then ...

- (Frobenius)

$$(u + v\sqrt{d})^p \equiv u + \left(\frac{d}{p}\right) v\sqrt{d} \bmod p.$$

- (AKS) $(X + b)^p \equiv X^p + b$ in $\mathbb{Z}[X]/\langle p, X^r - 1\rangle$.

Properties of primes III

If p is prime then ...

- (Pocklington) There is g such that $g^{p-1} \equiv 1 \bmod p$ and $g^{(p-1)/q} \not\equiv 1 \bmod p$ for any prime factor q of $p-1$.
- (ECPP) There is an elliptic curve E of order s with a point G such that $[s]G = O$ on E mod p and and $[s/q]G \neq O$ on E mod p for any prime factor q of s.
- (Adleman–Huang) There is an abelian variety A of order s with a point G such that $[s]G = O$ on A mod p and and $[s/q]G \neq O$ on A mod p for any prime factor q of s.

Speed

The criteria on the previous slides can be tested in time *polynomial* in $\log p$.

Probable primes and pseudoprimes

If **C** is one of these properties, or criteria, then call a p a **C**-probable prime if p has property **C**, or passes the **C** test.

Naturally every prime is a **C**-*probable prime*.

A composite number which passes the **C** test is a **C**-*pseudoprime*.

An example of a Fermat pseudoprime base 2 is $p = 341 = 11 \cdot 31$:

$$2^{340} \equiv 1 \bmod 341.$$

A Bayesian calculation I

- The proportion of bases b for which an odd composite n passes the Euler–Jacobi test is less than $\frac{1}{2}$.
- The proportion of bases b for which an odd composite n passes the Miller–Rabin (strong) test is less than $\frac{1}{4}$.

The probability that an odd number chosen uniformly at random in some interval is composite given that it passes one randomly chosen MR test is less than $\frac{1}{4}$.

A Bayesian calculation II

Let $p(k,t)$ denote the probability that an odd k-bit integer chosen uniformly at random is composite given that it passes k random MR tests.
For $k \geq 100$ and $5 \leq t \leq k/9 + 2$,

$$p(k,t) \leq 0.4\, k 2^t \left(0.6 \cdot 2^{-2\sqrt{k(t-2)}} + 2^{-t\sqrt{k/2}}\right);$$

and for $t > k/9 + 2$,

$$p(k,t) \leq 0.4\, k \left(11.32\sqrt{k}\, 2^{-2t-k/3} + 2^{t-t\sqrt{k/2}}\right).$$

For $t = 6$ and $k = 250$ this is less than 2^{-56} and for $t = 10$ and $k = 2000$ the probability of a wrong answer is less than 2^{-228}.

Similar results hold for quadratic tests such as Frobenius.

Distribution of primes

Let $\pi(x)$ denote the number of primes up to x.

- (PNT) $\pi(x) \sim \frac{x}{\log x}$
- (PNT) $\pi(x) = \mathrm{Li}(x) + \text{error}$

The *Riemann hypothesis* is equivalent to the error term being $O\left(x^{1/2+\epsilon}\right)$.

To what extent may we regard the primes as a *random* sequence with local density $1/\log x$?

“Properties” of primes

The density of primes would suggest that:

- (Goldbach conjecture) Every even number is the sum of two primes
- (Hardy-Littlewood) Infinitely many prime pairs $p, p+2$ or $p, 2p+1$, triples $p, 2p-1, 3p-2$, etc.
- The factors of $p-1$ should look like the factors of a general number of the same size

Sieve methods have made progress against the problems. For example, every sufficiently large even number is a sum of a prime and a product of at most two primes.

Carmichael numbers

A *Carmichael number* or *absolute pseudoprime* is a Fermat pseudoprime to every possible base.
Examples:

$$561 = 3 \cdot 11 \cdot 17$$

$$1729 = 7 \cdot 13 \cdot 19.$$

The numbers 2432870015941 and 3290922560713061 of the first slide are Carmichael numbers.

Let $\lambda(n)$ denote the exponent of the multiplicative group $(\mathbb{Z}/n)^*$.
A Carmichael number has $\lambda(n)|n-1$. Indeed,
$\lambda(561) = \text{lcm}\{2, 10, 16\} = 80$.

A composite number n is a Carmichael number iff it is odd, square-free and $p|n \Rightarrow p-1|n-1$.

Distribution of pseudoprimes

Let $P(x)$ denote the number of Fermat pseudoprimes base 2 and $C(x)$ the number of Carmichael numbers up to x.

- $C(X) \leq P(x) \ll x \cdot \exp\left(-\frac{\log x \log\log\log x}{\log\log x}\right)$

X	$P(X)$	$EP(X)$	$SP(X)$	$C(X)$
10^4	22	12	5	7
10^5	78	36	16	16
10^6	245	114	46	43
10^7	750	375	162	105
10^8	2057	1071	488	255
10^9	5597	2939	1282	646
10^{10}	14884	7706	3291	1547
25.10^9	21853	11347	4842	2163
10^{11}	38975	20417	8607	3605
10^{12}	101629	53332	22412	8241
10^{13}	264239	124882	58897	19279

Counts of Carmichael numbers

Values of $C(X)$ for X in powers of 10 up to 10^{21}.

x	$C\left(10^x\right)$
3	1
4	7
5	16
6	43
7	105
8	255
9	646
10	1547
11	3605
12	8241
13	19279
14	44706
15	105212
16	246683
17	585355
18	1401644
19	3381806
20	8220777
21	20138200

Conjectural

If $6k+1$, $12k+1$, $18k+1$ are all prime then

$$N = (6k+1)(12k+1)(18k+1)$$

is a Carmichael number.

Note that in this case $\lambda(N) = 36k$.

Example:

$$1729 = 7 \cdot 13 \cdot 19.$$

If the Hardy–Littlewood prime tuples conjecture is true, then there are infinitely many Carmichael numbers of this form alone.

There are at least three Carmichael numbers

Set $L = 120$ and put

$$S = \{p : p - 1 | L,\ p \not| L\} = \{7, 11, 13, 31, 41, 61\}.$$

Form all products n of elements of S such that $n \equiv 1 \bmod L$:

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

$$7 \cdot 13 \cdot 31 \cdot 61 = 172081$$

$$11 \cdot 13 \cdot 41 \cdot 61 = 852841.$$

For each such n we have

$$p | n \Rightarrow p - 1 | L | n - 1.$$

There are at least 2^{128} Carmichael numbers

Set $L = 23284800 = 2^6 \cdot 3^3 \cdot 5^2 \cdot 7^2 \cdot 11$ and put

$$S = \{p : p - 1 | L,\ p \not| L\} = \{13, 17, 19, \ldots, 3880801\}$$

There are 155 primes in S. Note that the first 27 have the property that the 2^{27} products cover $(\mathbb{Z}/L)^*$. Hence all possible products of the remaining 128 elements of S can be suitably "adjusted" to form an n such that $n \equiv 1 \bmod L$: there are at least 2^{128} such n, all satisfying

$$p|n \Rightarrow p - 1|L|n - 1.$$

The Erdős heuristic

Fix a parameter x and m. Let $k = \log x/m \log\log x$.
Set L to be the least common multiple of the integers up to $\log x/\log\log x$: we have $L = x^{o(1)}$. Set

$$S = \{\log x < p < (\log x)^m : p - 1|L\}.$$

Let T be the set of all products of distinct elements of S: the size of T is at least $x^{1-1/m}$ and the typical element of T is of size $X = x^{\log x}$. The number of elements of T which are congruent to 1 mod L should be about $|T|/L$.
If so, then we have

$$C(X) > X \cdot \exp\left(-\frac{\log X \log\log\log X}{\log\log x}\right).$$

There are infinitely many Carmichael numbers

Theorem

(Alford, Granville, Pomerance)

$$C(X) \gg X^{2/7}$$

(Harman)

$$C(X) \gg X^{0.332}$$

Ingredients

Added ingredients:

- An estimate for the size of a set whose partial products cover an abelian group
  - If G is an Abelian group of order n and exponent m then the products of a set of size $m + m\log(n/m)$ cover G.
- A bound for the number of primes p with $p - 1$ sufficiently smooth (free of large prime factors)
  - The number of primes up to x free of prime factors $> x^y$ is $> \gamma_y \pi(x)$; valid for $y > 1/2\sqrt{e}$
- An estimate for the distribution of primes in most arithmetic progressions of small difference
  - The number of primes in an arithmetic progression with difference d is $> \frac{1}{2}\pi(x)/\phi(d)$ for almost all $d < x^B$; valid for $B < \frac{5}{12}$

It is natural to conjecture that

$$C(X) \sim X \cdot \exp\left(-\frac{\log X \log\log\log X}{\log\log X}\right)$$

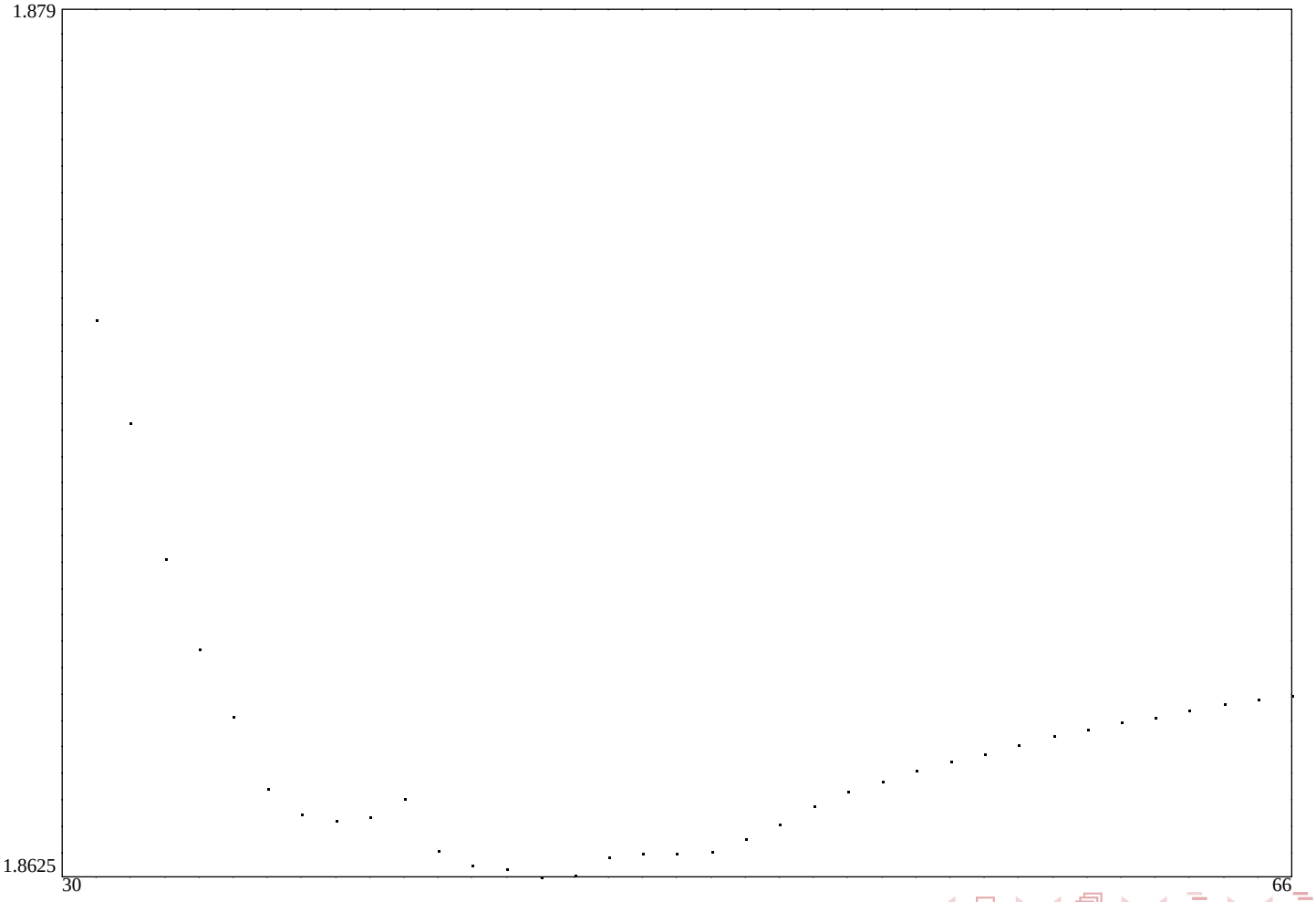
and more precisely that if we define $k(X)$ by

$$C(X) = X \cdot \exp\left(-k(X)\frac{\log X \log\log\log X}{\log\log X}\right)$$

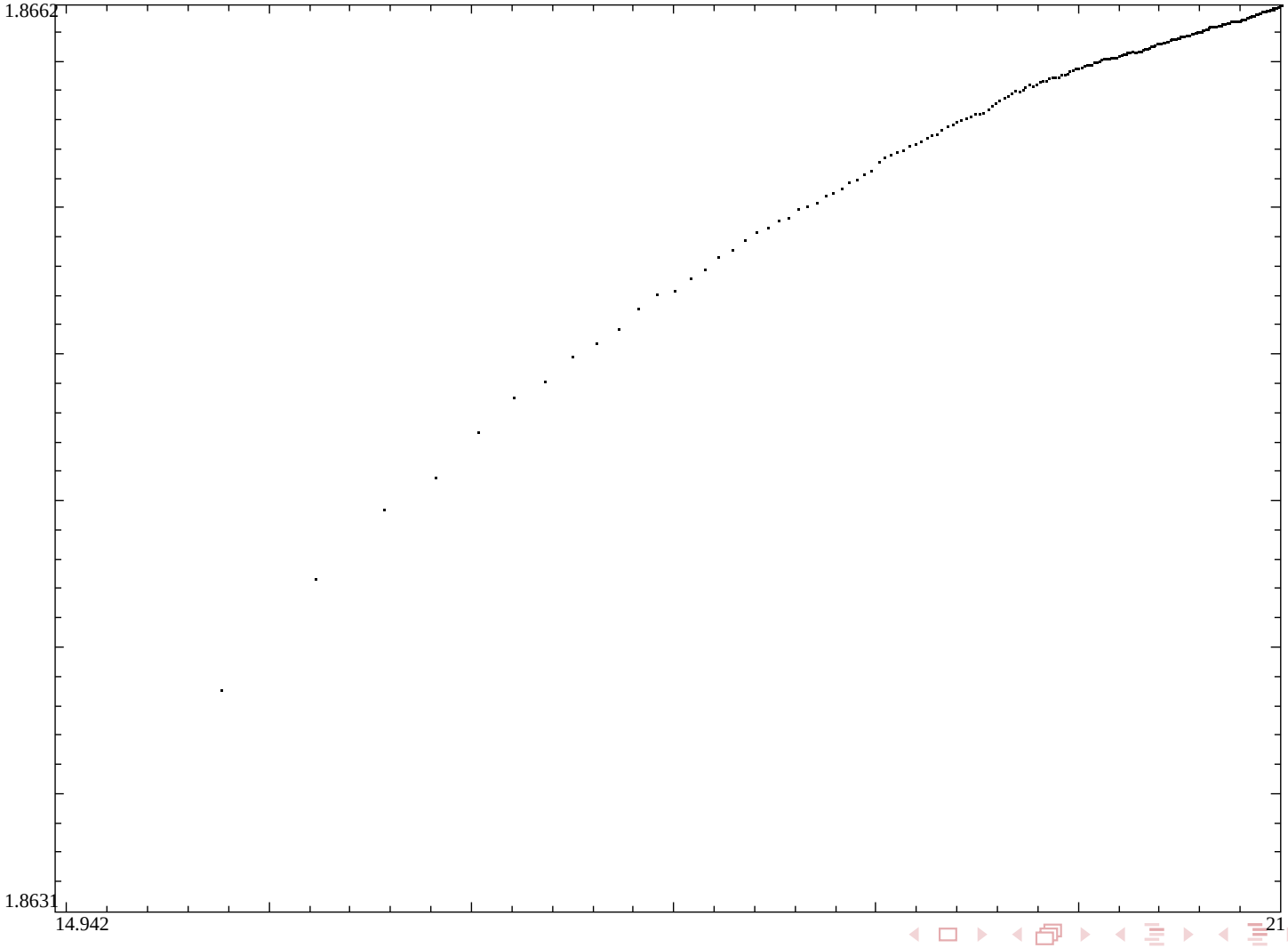
then

$$k(X) \to 1$$

$k(X)$ versus $\log X$



$k(X)$ versus $\log X$



It seems likely that the conjecture $k \to 1$ is too precise; the function $\log X \log\log\log X / \log\log X$ is only a convenient placeholder for a function implicitly defined within the heuristic argument.

The class P

We are interested in the complexity of algorithms for determining membership of sets of natural numbers (always encoded as binary strings in a natural way).

A set, or property, S is in class **P** if there is an algorithm which determines membership of S for n-bit numbers in time polynomial in n: that is, $O(n^c)$ for some constant c.

PERFECTPOWERS, the set of perfect powers, is in P.

The class NP

A set, or property, S is in class **NP** if there is an algorithm which determines membership of S for n-bit numbers in time polynomial in n given the correct auxiliary input ("witness" or "certificate").

COMPOSITES, the set of composite numbers, is in NP: for witness we may take a factor of the input number.

The class RP

A set, or property, S is in class **RP** if there is an algorithm which determines membership of S for n-bit numbers in time polynomial in n given the correct auxiliary input ("witness" or "certificate"), and a positive proportion of witnesses are valid.

A set is in **co-P** or **co-NP** if the complementary set is in **P** or **NP** respectively.

PRIMES and COMPOSITES

- PRIMES is in co-NP (COMPOSITES is in NP)
  - Certificate: a factor.
- PRIMES is in co-RP (COMPOSITES is in RP)
  - Certificate: a base failing the Miller–Rabin (“strong”) test.
- PRIMES is in NP
  - Certificate: a primitive root modulo p, together with the prime factors of $p - 1$ and a recursive certificate of primality of those factors.
- PRIMES is in RP
  - Certificate: an abelian variety with smooth order modulo p.

PRIMES is in P: I

- Theorem. If the Generalised Riemann Hypothesis (GRH) is true, then any consecutive $2(\log n)^2$ numbers generate the multiplicative group modulo n.
- Without GRH the best we can say is $n^{1/2\sqrt{e}}$.
- Corollary. GRH $\Rightarrow$ PRIMES is in P.

PRIMES is in P: II

Theorem

PRIMES is in P

- If the order of p modulo r is greater than $(\log p)^2$ and the AKS condition

$$(X + a)^p \equiv X^p + a \text{ in } \mathbb{Z}[X]/\langle p, X^r - 1\rangle$$

  is satisfied for all $a \leq \sqrt{r} \log p$, then p is prime.

- Such an r exists and is less than $(\log p)^3$
- The primality of p may be determined in time $O((\log p)^{15/2})$.

PRIMES is in P: III

Ingredients:

- An estimate for the size of a set whose partial products cover an abelian group

- A result stating that for primes q, the largest prime factor of $q-1$ behaves sufficiently like that of a "random" integer of the same size.
  - The number of primes q up to x such that $q-1$ has largest prime factor $> x^u$ is $> c_u \pi(x)$, valid for $u \leq \frac{2}{3}$.

Coda

- Experimental results fuel theoretical investigation
- Explicit arithmetic can simulate random phenomena
- The proofs lie deep (sometimes too deep)
- Wir mussen wissen — wir werden wissen