Heuristics for Prime Statistics
Brown Univ.
Feb. 11, 2006

K. Conrad, UConn

Two quotes about prime numbers

Mathematicians have tried in vain to this day to discover some order in the sequence of prime numbers, and we have reason to believe that it is a mystery into which the human mind will never penetrate.
L. Euler (1751)

Although the prime numbers are rigidly determined, they somehow feel like experimental data.
T. Gowers (2002)

Lists of primes

There are infinitely many of them:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, \ldots.$$

But are there infinitely many primes of special forms?

- (Mersenne) $2^n - 1$: $3, 7, 31, 127, 8191, \ldots$? (43rd example was found on Dec. 15, 2005.)
- (Fermat) $2^{2^n} + 1$: $3, 5, 17, 257, 65537$. No more known.
- (Dirichlet) $20n + 9$: $29, 89, 109, 149, 229, \ldots$.
- (Euler) $n^2 + 1$: $2, 5, 17, 37, 101, \ldots$?
- twin primes: $(3, 5), (5, 7), (11, 13), (17, 19), (29, 31), \ldots$?

Relevance of Mersenne and Fermat primes

A *perfect number* is an integer equal to the sum of its proper factors:

$$6 = 1 + 2 + 3, \quad 28 = 1 + 2 + 4 + 7 + 14.$$

Theorem. (Euler) The even perfect numbers are the numbers $2^{n-1}(2^n - 1)$, where $2^n - 1$ is prime.

Odd perfect numbers are not expected to exist.

Visit `http://www.mersenne.org/` and help find more Mersenne primes.

Theorem. (Gauss) A regular polygon with d sides can be constructed using an unmarked straightedge and a compass if and only if $d = 2^r p_1 p_2 \dots p_m$, where $r \geq 0$ and the p_i's are distinct Fermat primes.

Example. Using an unmarked straightedge and compass, a regular 9-gon can't be constructed and a regular 17-gon can be constructed.

Open questions about prime values of polynomials

- Show there are infinitely many primes of the form $n^2 + 1$.
- For any polynomial $f(T)$, decide when $f(n)$ should be prime for infinitely many integers n and quantify this.
- Show there are infinitely many twin primes.
- For two polynomials $f(T)$ and $g(T)$, decide when $f(n)$ and $g(n)$ are simultaneously prime for infinitely many n and quantify this.

We are *not* looking for a formula for all primes or a formula whose values are *only* primes. (A non-constant polynomial will not take only prime values.) We are looking at prime values of polynomials, asking whether prime values occur infinitely often and, if so, how often.

Why does this matter?

1) Mallory's answer: it's there.

2) Decimal expansions: the decimal expansion of any fraction is eventually periodic.

$$\begin{aligned}
\frac{1}{3} &= .33333333333333333333333\ldots \\
\frac{1}{6} &= .16666666666666666666666\ldots \\
\frac{1}{7} &= .\underbrace{142857}_{\text{period } 6}142857142857142857\ldots \\
\frac{1}{17} &= .\underbrace{0588235294117647}_{\text{period } 16}0588235\ldots
\end{aligned}$$

For prime p, the decimal expansion of $1/p$ has period length dividing $p-1$.

Example. $7, 17, 19, 23, 29, 47, 59, 61, 97, \ldots$?

Is the period length $p-1$ infinitely often? If $20n+9$ and $40n+19$ are simultaneously prime infinitely often then the answer is YES. (Other explanations are possible.)

Why does this matter?

3) Finding Pythagorean triples ($a^2 + b^2 = c^2$) is the same as solving $x^2 + y^2 = 1$ in fractions (*e.g.*, $(3/5)^2 + (4/5)^2 = 1$). Is there a solution *in fractions* to "quadratic equations" like $x^2 + 3y^2 - 10z^2 = 14$? There is an algorithm to settle this question, but the proof that it works depends on knowing certain polynomials are prime infinitely often.

4) In 1994, Thomas Nicely discovered a bug in the Pentium processor's floating point division. For instance, the Pentium chip produced the following results:

$$\frac{5505001}{294911}\text{“}=\text{”}18.66600093 \quad \text{(it's really 18.66665197).}$$

$$4195835 - \left(\frac{4195835}{3145727}\right) 3145727\text{“}=\text{”}256.$$

It was a public relations nightmare for Intel.

How did Nicely find the bug? He wasn't searching for it, but was studying problems related to the frequency of twin primes. The Pentium chip incorrectly computed the reciprocals of the twin primes 824633702441 and 824633702443.

Asymptotic estimates

Set

$$\pi(x) = \#\{n \leq x : n \text{ is prime}\}.$$

It is unreasonable to ask for an exact formula for $\pi(x)$, but there is an elementary *asymptotic* formula for $\pi(x)$.

Call two (eventually) positive functions $f(x)$ and $g(x)$ *asymptotic* when

$$\frac{f(x)}{g(x)} \to 1$$

as $x \to \infty$. We then write $f(x) \sim g(x)$.

Examples. $x^2 + 7x \sim x^2, \quad \sqrt{4x^2 + 3x} + \sin x \sim 2x.$

Exercise. If $f(x) \sim g(x)$ and $\sum_{n \geq 1} f(n) = \infty$ then

$$\sum_{n \leq x} f(n) \sim \sum_{n \leq x} g(n).$$

Example. Since $\displaystyle\int_n^{n+1} t^2 \mathrm{d}t = n^2 + n + \frac{1}{3} \sim n^2,$

$$\sum_{n \leq x} n^2 \sim \sum_{n \leq x} \int_n^{n+1} t^2 \mathrm{d}t = \int_1^{[x]+1} t^2 \mathrm{d}t \sim \frac{([x]+1)^3}{3} \sim \frac{x^3}{3}.$$

Prime number theorem

Theorem. (Hadamard, de la Vallée Poussin, 1896)

$$\pi(x) \sim \frac{x}{\log x},$$

where log is the natural logarithm.

Since $\dfrac{x}{\log x} \sim \displaystyle\int_2^x \frac{\mathrm{d}t}{\log t}$ and $\displaystyle\int_n^{n+1} \frac{\mathrm{d}t}{\log t} \sim \frac{1}{\log n}$, the prime number theorem is equivalent to

$$\pi(x) \sim \sum_{2 \le n \le x} \frac{1}{\log n}.$$

x	10^4	10^5	10^6	10^7	10^8
$\pi(x)$	1229	9592	78498	664579	5761455
$\dfrac{x}{\log x}$	1085	8685	72382	620420	5428681
Ratio	1.131	1.104	1.084	1.071	1.061
$\pi(x)$	1229	9592	78498	664579	5761455
$\displaystyle\sum_{2 \le n \le x} \frac{1}{\log n}$	1245	9629	78627	664918	5762209
Ratio	.9863	.9960	.9983	.9994	.9998

Counting and comparing

Set

$$\begin{aligned} \pi_{T^2+1}(x) &= \#\{n \le x : n^2+1 \text{ is prime}\}, \\ \pi_{\text{twin}}(x) &= \#\{n \le x : n \text{ and } n+2 \text{ are prime}\}. \end{aligned}$$

x	10^4	10^5	10^6	10^7	10^8
$\pi(x)$	1229	9592	78498	664579	5761455
$\pi_{T^2+1}(x)$	841	6656	54110	456362	3954181
$\pi_{\text{twin}}(x)$	205	1224	8169	58980	440312
$\frac{\pi_{T^2+1}(x)}{\pi(x)}$	.6842	.6939	.6893	.6866	.6863
$\frac{\pi_{\text{twin}}(x)}{\pi(x)}$	.1668	.1276	.1040	.0887	.0764

- Is $\pi_{T^2+1}(x) \sim K\pi(x)$ for a constant $K \approx .68$?
- It appears that $\pi_{\text{twin}}(x)$ has a slower order of growth than $\pi(x)$. How much slower?

A probabilistic viewpoint

The prime number theorem says

$$\pi(x) \sim \frac{x}{\log x} \sim \sum_{2 \le n \le x} \frac{1}{\log n}.$$

If we heuristically set

$$\text{Prob}(n \text{ prime}) = \frac{1}{\log n}$$

then the number of primes up to x can be counted asymptotically by "adding up the probabilities."

> *Probability* is not a notion of mathematics, but of philosophy or physics.
> G. H. Hardy, J. E. Littlewood (1923)

From asymptotics to probability

The provable estimates

$$\#\{n \le x : n \text{ is prime}\} \sim \frac{x}{\log x} \sim \sum_{2 \le n \le x} \frac{1}{\log n},$$

$$\#\{n \le x : n \text{ is even}\} \sim \frac{x}{2} \sim \sum_{n \le x} \frac{1}{2},$$

$$\#\{n \le x : n = \square\} \sim \sqrt{x} \sim \sum_{n \le x} \frac{1}{2\sqrt{n}}$$

suggest the heuristics

$$\begin{aligned} \text{Prob}(n \text{ is prime}) &= \frac{1}{\log n}, \\ \text{Prob}(n \text{ is even}) &= \frac{1}{2}, \\ \text{Prob}(n = \square) &= \frac{1}{2\sqrt{n}}. \end{aligned}$$

Estimating $\pi_{T^2+1}(x)$ with probabilities

If we believe

$$\text{Prob}(n \text{ is prime}) = \frac{1}{\log n}$$

then it seems reasonable to set

$$\text{Prob}(n^2+1 \text{ is prime}) = \frac{1}{\log(n^2+1)},$$

so we "expect"

$$\begin{aligned}
\pi_{T^2+1}(x) &= \#\{n \le x : n^2+1 \text{ is prime}\} \\
\text{“}=\text{”} &\sum_{n \le x} \text{Prob}(n^2+1 \text{ is prime}) \\
&= \sum_{n \le x} \frac{1}{\log(n^2+1)} \\
&\sim \sum_{2 \le n \le x} \frac{1}{2\log n} \\
&\sim \frac{1}{2}\frac{x}{\log x}.
\end{aligned}$$

But our data suggested $\pi_{T^2+1}(x) \sim C\dfrac{x}{\log x}$ where $C \approx .68$. Maybe we would find $C \approx .5$ if we waited longer?

Fixing the prime heuristic, I

When counting $\pi_{T^2+1}(x) = \#\{n \leq x : n^2+1 \text{ is prime}\}$, we should take into account that we are looking *only* at numbers of the form n^2+1. The heuristic formula

$$\text{Prob}(n^2+1 \text{ is prime}) = \frac{1}{\log(n^2+1)}$$

came from thinking of the numbers n^2+1 as part of the whole set of positive integers. But integers of the form n^2+1, considered collectively, have different divisibility properties than all integers. These numbers begin as

$$2, 5, 10, 17, 26, 37, 50, 65, 82, 101, 122, 145, 170, \ldots.$$

They alternate even and odd, like all integers, but where are the multiples of 3? There are none:

$$\begin{aligned} (3k)^2+1 &= 3(3k^2)+1, \\ (3k+1)^2+1 &= 3(3k^2+2k)+2, \\ (3k+2)^2+1 &= 3(3k^2+4k+1)+2. \end{aligned}$$

Fixing the prime heuristic, II

A random integer n has "probability" $1/5$ of being a multiple of 5, but the sequence of numbers n^2+1 contains multiples of 5 twice as often:

$$2, \mathbf{5}, \mathbf{10}, 17, 26, 37, \mathbf{50}, \mathbf{65}, 82, 101, 122, \mathbf{145}, \mathbf{170}, \ldots.$$

So n^2+1 is "more likely" to be prime than n due to the effect of 3 and "less likely" to be prime than n due to the effect of 5. The effect of 2 on primality of n and n^2+1 is "the same." Let's quantify these "local" effects coming from divisibility of numbers by 2, by 3, by 5, and so on.

Heuristic: $\text{Prob}(n^2+1 \text{ prime}) = \dfrac{C}{\log(n^2+1)}$,

where C is a constant that accounts for divisibility features of the sequence n^2+1 by comparison with divisibility features of all integers n ("conditional probability"). What is C???

Determining the constant C

$$\text{Prob}(n^2+1 \text{ prime}) = \frac{C}{\log(n^2+1)}$$

For each prime p, whether or not n^2+1 is divisible by p only depends on whether or not r^2+1 is divisible by p, where r is the remainder when n is divided by p.

Example. $3^2+1=10$ is divisible by 5 and $(5k+3)^2+1=5(5k^2+6k)+10$ is divisible by 5.

Example. $1^2+1=2$ is not divisible by 5 and $(5k+1)^2+1=5(5k^2+2k)+2$ is not divisible by 5.

For each prime p, let $\omega(p)$ be the number of remainders $0 \le r \le p-1$ such that r^2+1 is divisible by p.

Example. At 2, $0^2+1=1$ and $1^2+1=2$, so $\omega(2)=1$.

$\omega(2)=1, \omega(3)=0, \omega(5)=2, \omega(7)=0, \omega(11)=0, \omega(13)=2$

The "probability" that n is not divisible by p is $1-1/p$ and the "probability" that n^2+1 is not divisible by p is $1-\omega(p)/p$. Set

$$C = \prod_p \left(\frac{1-\omega(p)/p}{1-1/p} \right) = \frac{1-1/2}{1-1/2} \cdot \frac{1}{1-1/3} \cdot \frac{1-2/5}{1-1/5} \cdots .$$

A conjecture for primality of n^2+1

It is proposed that

$$\pi_{T^2+1}(x) \sim \sum_{n \le x} \frac{C}{\log(n^2+1)},$$

or in simpler terms

$$\pi_{T^2+1}(x) \sim \frac{C}{2} \frac{x}{\log x},$$

where

$$C = \prod_p \frac{1-\omega(p)/p}{1-1/p} \approx 1.37281346, \quad \frac{C}{2} \approx .6864067.$$

In the table, "Approx." comes from the summation up to x.

x	$\pi_{T^2+1}(x)$	Approx.	Ratio
10^2	19	22.4	.849
10^3	112	123.6	.906
10^4	841	857.0	.981
10^5	6656	6611.6	1.007
10^6	54110	53972.1	1.003
10^7	456362	456406.1	1.000

General conjecture

Let $f(T)$ be an irreducible polynomial with integer coefficients. Set

$$\pi_f(x) = \#\{n \le x : |f(n)| \text{ is prime}\}$$

and for primes p set

$$\omega_f(p) = \#\{0 \le r \le p-1 : f(r) \text{ is divisible by } p\}.$$

Heuristic: $\text{Prob}(|f(n)| \text{ is prime}) = \dfrac{C(f)}{\log|f(n)|}$,

where $C(f)$ accounts for non-divisibility features of the sequence of numbers $f(n)$:

$$C(f) = \prod_p \frac{1-\omega_f(p)/p}{1-1/p}.$$

Conjecture. (Hardy–Littlewood, Bateman–Horn) *For an irreducible polynomial $f(T)$ with $C(f) > 0$,*

$$\pi_f(x) \overset{?}{\sim} \sum_{n\le x}{}' \frac{C(f)}{\log|f(n)|} \sim \frac{C(f)}{\deg f}\frac{x}{\log x}.$$

What is known?

$$\pi_f(x) \stackrel{?}{\sim} \frac{C(f)}{\deg f}\frac{x}{\log x}, \quad C(f) = \prod_p \frac{1-\omega_f(p)/p}{1-1/p}$$

- This conjecture is a theorem (of Dirichlet) when $\deg f = 1$.
- There are *no* examples with $\deg f > 1$ for which it is proved that $\pi_f(x) \to \infty$.
- **Exercise**. Show $\pi_f(x) \sim \dfrac{C}{\deg f}\dfrac{x}{\log x}$ is equivalent to

$$n\text{-th prime value of } f \sim \frac{\deg f}{C} n \log n.$$

- For irreducible $f(T)$, $C(f) = 0$ only when there is a prime p for which $f(n)$ is divisible by p for all n (*e.g.*, $f(T) = T^2 - T + 2$ and $p = 2$: $n^2 - n + 2$ is always even).

So we believe that when $f(T)$ is irreducible, $f(n)$ is prime for infinitely many n except in the "trivial" situation that all the numbers $f(n)$ have a common prime factor.

Simultaneous primality

Let $f(T)$ and $g(T)$ be irred. with integer coefficients. Set

$$\pi_{f,g}(x) = \#\{n \le x : |f(n)| \text{ and } |g(n)| \text{ are both prime}\}.$$

For "independently" chosen positive integers m and n,

$$\text{Prob}(m \text{ and } n \text{ are prime}) = \frac{1}{\log m} \cdot \frac{1}{\log n}.$$

How often are $f(n)$ and $g(n)$ prime (same n)? Set

$$\omega_{f,g}(p) = \#\{0 \le r \le p-1 : f(r) \text{ or } g(r) \text{ is div. by } p\},$$

$$C(f,g) = \prod_p \frac{1 - \omega_{f,g}(p)/p}{(1-1/p)^2}.$$

Then our heuristic for simultaneous primality is

$$\text{Prob}(f(n) \text{ and } g(n) \text{ are prime}) = \frac{C(f,g)}{\log|f(n)| \log|g(n)|}.$$

Conjecture. *If* $C(f,g) > 0$,

$$\pi_{f,g}(x) \stackrel{?}{\sim} \sum_{n \le x}{}' \frac{C(f,g)}{\log|f(n)| \log|g(n)|} \sim \frac{C(f,g)}{(\deg f)(\deg g)} \frac{x}{(\log x)^2}.$$

Examples

Example. Twin primes ($f(T) = T$, $g(T) = T + 2$). Then

$$C_{\text{twin}} = C(T, T+2) = 2 \prod_{p \neq 2} \frac{1 - 2/p}{(1 - 1/p)^2} \approx 1.32032363.$$

Conjecture. (Hardy–Littlewood, 1923).

$$\pi_{\text{twin}}(x) \stackrel{?}{\sim} \sum_{n \leq x} \frac{C_{\text{twin}}}{(\log n)(\log(n+2))} \sim C_{\text{twin}} \frac{x}{(\log x)^2}.$$

This would explain our numerical observation that $\pi_{\text{twin}}(x)$ grows at a slower order than $\pi_{T^2+1}(x) \stackrel{?}{\sim} C(T^2+1)\dfrac{x}{\log x}$.

Example. Prime pairs n and $n + 6$. Factors in $C(T, T+6)$ and C_{twin} are equal except at $p = 3$: $C(T, T+6) = 2C_{\text{twin}}$. So we expect $\pi_{T,T+6}(x) \sim 2\pi_{\text{twin}}(x)$.

x	10^4	10^5	10^6	10^7	10^8
$\pi_{\text{twin}}(x)$	205	1224	8169	58980	440312
$2\pi_{\text{twin}}(x)$	410	2448	16338	117960	880624
$\pi_{T,T+6}(x)$	411	2447	16386	117207	879980

Multiple primality

The probabilistic heuristic can be extended to predict the frequency of simultaneous primality of more than two irreducible polynomials.

Example. No "triple primes" $n, n+2, n+4$ for $n > 3$ since at least one is a multiple of 3!

Example. Pick an integer $d \geq 1$. Consider polynomials $T + a_1, T + a_2, \ldots, T + a_d$ where the a_i's are different positive integers divisible by no prime up to d. Then the number of n up to x such that $n + a_1, \ldots, n + a_d$ are all prime should be $\sim C\dfrac{x}{(\log x)^d}$ for some constant $C > 0$ depending only on the a_i's. In particular, this predicts there are arithmetic progressions of prime numbers with arbitrarily long length.

The longest known arithmetic progression of prime numbers has length 23:

$$44546738095860n + 56211383760397$$

for $0 \leq n \leq 22$ (Frind, Jobling, Underwood, 2004).

Theorem. (Green, Tao, 2004). *There are arbitrarily long arithmetic progressions of prime numbers.*

A final crazy case

Example. T and $T+1$.

There are not infinitely many prime pairs n and $n+1$: at least one of them is even! The constant $C(T, T+1)$ is 0 too.

But the *only* obstruction is caused by 2. Set

$$n_{\text{odd}} = \text{ odd part of } n.$$

n	1	2	3	4	5	6	7	8	9	10
n_{odd}	1	1	3	1	5	3	7	1	9	5
$(n+1)_{\text{odd}}$	1	3	1	5	3	7	1	9	5	11

Are there infinitely many n such that n_{odd} and $(n+1)_{\text{odd}}$ are both prime?

Exercise. Formulate a heuristic probability

$$\text{Prob}(n_{\text{odd}} \text{ is prime}) = \frac{C}{\log n_{\text{odd}}}$$

for a suitably defined constant C (hint: $C \neq 1$), extend it to simultaneous primality of n_{odd} and $(n+1)_{\text{odd}}$, and make a conjecture for how often n_{odd} and $(n+1)_{\text{odd}}$ are both prime.

References

S. Lang, "Math Talks for Undergraduates," Springer-Verlag, New York, 1999; Chap. 1

P. Ribenboim, "The New Book of Prime Number Records," Springer-Verlag, New York, 1996; Chap. 6

H. Riesel, "Prime Numbers and Computer Methods for Factorization," 2nd ed., Birkhauser, Boston, 1994; Chap. 3

W. Narkiewicz, "The Development of Prime Number Theory: from Euclid to Hardy and Littlewood," Springer-Verlag, Berlin, 2000; Sect. 6.7