A formally verified proof of the prime number theorem

Jeremy Avigad

Department of Philosophy

Carnegie Mellon University

http://www.andrew.cmu.edu/∼avigad

The prime number theorem

Let $\pi(x)$ denote the number of primes less than or equal to x.

The prime number theorem: $\pi(x)/x$ is asymptotic to $1/\ln x$, i.e.

$$\lim_{x \to \infty} \pi(x) \ln x / x = 1.$$

Conjectured by Gauss and Legendre, on the basis of computation, around 1800; proved by Hadamard and de la Vallée Poussin in 1896.

Kevin Donnelly, David Gray, Paul Raff, and I used Isabelle to verify:

$$(\lambda x.\ pi\ x * ln\ (real\ x)\ /\ (real\ x)) \longrightarrow 1$$

Outline

- Historical background
- Overview of the Selberg proof
- Overview of the formalization
- Interesting aspects of the formalization
  - Asymptotic reasoning
  - Calculations with reals
  - Casts between natural numbers, integers, and reals
  - Combinatorial reasoning with sums
  - Elementary workarounds
- Heuristic procedures for the reals

Chebyshev's advances (∼1850)

$$\theta(x) = \sum_{p \le x} \ln p$$

$$\psi(x) = \sum_{p^a \le x} \ln p = \sum_{n \le x} \Lambda(n) \quad \text{where}$$

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n = p^a, \text{ for some } a \ge 1 \\ 0 & \text{otherwise.} \end{cases}$$

- The prime number is equivalent to the statements $\lim_{x\to\infty} \theta(x)/x = 1$ and $\lim_{x\to\infty} \psi(x)/x = 1$.
- For x large enough,

$$0.92 < \pi(x) \ln x/x < 1.11.$$

More history

In 1859, Riemann introduces the complex-valued function, ζ.

In 1894, von Mangoldt reduced the PNT to showing that ζ has no roots with real part equal to 1.

This was done by Hadamard and de la Vallée Poussin, independently, in 1896.

In 1921, Hardy expressed doubts that there is a proof that does not essentially use these ideas.

In 1948, Selberg and Erdös found elementary proofs based on Selberg's "symmetry formula."

Outline

- Historical background
- Overview of the Selberg proof
- Overview of the formalization
- Interesting aspects of the formalization
  - Asymptotic reasoning
  - Calculations with reals
  - Casts between natural numbers, integers, and reals
  - Combinatorial reasoning with sums
  - Elementary workarounds
- Heuristic procedures for the reals

Asymptotic reasoning

View $\pi(x)$ as a step function from $\mathbb{R}$ to $\mathbb{R}$.

Analytic number theory provides a toolbox for characterizing growth rates.

For example, $f = O(g)$ means: there is a constant, C, such that for every x,

$$|f(x)| \leq C|g(x)|.$$

Sometimes, one really means "for all but a few exceptional cases of x," or "for large enough x."

Examples

Here are some identities involving ln:

$$\ln(1+1/n) = 1/n + O(1/n^2)$$

$$\sum_{n \le x} 1/n = \ln x + O(1)$$

$$\sum_{n \le x} \ln n = x \ln x - x + O(\ln x)$$

$$\sum_{n \le x} \ln n/n = \ln^2 x/2 + O(1)$$

These, and a few others, form a starting point for the Selberg proof.

Chebyshev's results

Fairly direct calculations yield $\theta(x)/x \to 1$ and $\pi(x)\ln x/x \to 1$ from $\psi(x)/x \to 1$.

This allows us to prove the prime number theorem in the form $\psi(x)/x \to 1$.

Along the way, we need $\psi(x) = O(x)$.

There is a nice way to do this, using binomial coefficients.

Combinatorial tricks

Since $d \mapsto n/d$ permutes the set of divisors of n,

$$\sum_{d|n} f(d) = \sum_{d|n} f(n/d).$$

Enumerating pairs d, d' such that $dd' \leq n$ in two different ways yields

$$\sum_{d \leq n} \sum_{d' \leq n/d} f(d, d') = \sum_{dd' \leq n} f(d, d') = \sum_{c \leq n} \sum_{d|c} f(d, c/d).$$

A similar argument yields

$$\sum_{d|n} \sum_{d'|(n/d)} f(d, d') = \sum_{dd'|n} f(d, d') = \sum_{c|n} \sum_{d|c} f(d, c/d).$$

Combinatorial tricks

The following is a version of the "partial summation formula": if $a \leq b$, $F(n) = \sum_{i=1}^{n} f(i)$, and G is any function, then

$$\sum_{n=a}^{b} f(n+1)G(n+1) = F(b+1)G(b+1) - F(a)G(a+1) - \sum_{n=a}^{b-1} F(n+1)(G(n+2) - G(n+1)).$$

This is a discrete analogue of integration by parts.

It is easily verified by induction.

Euler's function μ

A positive natural number n is *square free* if $n = p_1 p_2 \cdots p_s$ with p_i's distinct.

$$\mu(n) = \begin{cases} (-1)^s & \text{if } n \text{ is squarefree and } s \text{ is as above} \\ 0 & \text{otherwise.} \end{cases}$$

A remarkably useful fact regarding μ is that for $n > 0$,

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise.} \end{cases}$$

This is clear for $n = 1$.

Euler's function μ

If $n = p_1^{j_1} p_2^{j_2} \cdots p_s^{j_s}$, define the *radical* of n to be $p_1 p_2 \cdots p_s$.

Then

$$\sum_{d|n} \mu(d) = \sum_{d|rad(n)} \mu(d)$$

$$= \sum_{d|rad(n),\, p_1|d} \mu(d) + \sum_{d|rad(n),\, p_1 \nmid d} \mu(d),$$

and the two terms cancel.

Möbius inversion

Suppose $f(n) = \sum_{d|n} g(d)$. Then

$$\begin{aligned}
\sum_{d|n} \mu(d) f(n/d) &= \sum_{d|n} \mu(d) \sum_{d'|(n/d)} g((n/d)/d') \\
&= \sum_{d|n} \sum_{d'|(n/d)} \mu(d) g((n/d)/d') \\
&= \sum_{c|n} \sum_{d|c} \mu(d) g(n/c) \\
&= \sum_{c|n} g(n/c) \sum_{d|c} \mu(d) \\
&= g(n),
\end{aligned}$$

expresses g in terms of f.

Selberg's formula

All these pieces come together in the proof of Selberg's symmetry formula:

$$\sum_{n \le x} \Lambda(n) \ln n + \sum_{n \le x} \sum_{d|n} \Lambda(d)\Lambda(n/d) = 2x \ln x + O(x).$$

There are many variants of this identity.

The reason it is useful is that there are two terms in the sum on the left, each sensitive to the presence of primes in different ways.

Selberg's proof involves cleverly balancing the two terms off each other, to show that in the long run, the density of the primes has the appropriate asymptotic behavior.

The error term

Let $R(x) = \psi(x) - x$ denote the "error term."

By Chebyshev's equivalences the prime number theorem amounts to the assertion $\lim_{x\to\infty} R(x)/x = 0$.

With some delicate calculation, the symmetry formula yields:

$$|R(x)| \ln^2 x \leq 2 \sum_{n\leq x} |R(x/n)| \ln n + O(x \ln x). \tag{1}$$

Selberg used this to show that, given a bound $|R(x)| \leq ax$ for sufficiently large x, one can get a better bound, $|R(x)| \leq a'x$, for sufficiently large x.

These bounds approach 0.

Outline

- Historical background
- Overview of the Selberg proof
- Overview of the formalization
- Interesting aspects of the formalization
  - Asymptotic reasoning
  - Calculations with reals
  - Casts between natural numbers, integers, and reals
  - Combinatorial reasoning with sums
  - Elementary workarounds
- Heuristic procedures for the reals

Overview of the formalization

To start with, we needed good supporing libraries:

- a theory of the natural numbers and integers, including properties of primes and divisibility, and the fundamental theorem of arithmetic
- a library for reasoning about finite sets, sums, and products
- a library for the real numbers, including properties of ln

More specific supporting libraries include:

- properties of the μ function, combinatorial identities, and variants of the Möbius inversion formula
- a library for asymptotic "big O" calculations
- a number of basic identities involving sums and ln
- Chebyshev's theorems

Overview of the formalization

Specific components of the Selberg proof are:

- the Selberg symmetry formula
- the inequality involving $R(n)$
- a long calculation to show $R(n)$ approaches 0

This outline is clearly discernible in the list of theory files, online at

http://www.andrew.cmu.edu/user/avigad/isabelle

Overview of the formalization

Here is a formulation of Möbius inversion:

$$\begin{aligned}&\mathit{ALL}\ n.\ (0 < n \longrightarrow\\ &\quad f\ n = (\sum\ d \mid d\ \mathit{dvd}\ n.\ g(n\ \mathit{div}\ d))) \Longrightarrow 0 < (n::\mathit{nat}) \Longrightarrow\\ &\quad g\ n = (\sum\ d \mid d\ \mathit{dvd}\ n.\ \mathit{of\text{-}int}(\mathit{mu}(\mathit{int}(d))) * f\ (n\ \mathit{div}\ d))\end{aligned}$$

Here is one of the identities given above:

$$\begin{aligned}&(\lambda x.\ \sum i{=}1..\mathit{natfloor}(\mathit{abs}\ x).\\ &\quad \mathit{ln}\ (\mathit{real}\ i)\ /\ (\mathit{real}\ i)) =o\\ &\quad (\lambda x.\ \mathit{ln}(\mathit{abs}\ x + 1)\hat{}2\ /\ 2) +o\ O(\lambda x.\ 1)\end{aligned}$$

Overview of the formalization

Here is a version of Selberg's symmetry formula:

$$(\lambda x.\ \sum n = 1..natfloor\ (abs\ x) + 1.$$
$$Lambda\ n * ln\ (real\ n)) + (\lambda x.\ \sum n{=}1..natfloor\ (abs\ x) + 1.$$
$$(\sum\ u \mid u\ dvd\ n.\ Lambda\ u * Lambda\ (n\ div\ u)))$$
$$=o\ (\lambda x.\ 2 * (abs\ x + 1) * ln\ (abs\ x + 1)) +o\ O(\lambda x.\ abs\ x + 1)$$

Finally, here is the error estimate provided above:

$$(\lambda x.\ abs\ (R\ (abs\ x + 1)) * ln\ (abs\ x + 1)\ \hat{}\ 2) <o$$
$$(\lambda x.\ 2 * (\sum\ n = 1..natfloor\ (abs\ x) + 1.$$
$$abs\ (R\ ((abs\ x + 1)\ /\ real\ n)) * ln\ (real\ n))) =o$$
$$O(\lambda x.\ (abs\ x + 1) * (1 + ln\ (abs\ x + 1)))$$

Overview of the formalization

There are at least three reasons not to provide too much detail:

- Our proof followed textbook presentations (due to Shapiro, Nathanson) closely.
- The proof scripts have not been polished, and so are not particularly nice.
- Much of it is not optimal; we know it is possible to do better.

Instead I will focus on:

- Details that diverge from the mathematical presentation.
- Novel features of the formalization.
- Areas where better support should be possible.

Overview of the formalization

Some statistics regarding length, and time, are given in the associated paper.

A lot of time and effort was spent:

- Building basic libraries of easy facts.
- Spelling out "straightforward" inferences.
- Finding the right lemmas and theorems to apply.
- Entering long formulas and expressions formally and correctly.

We suspect that these requirements will continue to diminish.

On a personal note, I am entirely convinced that formal verification of mathematics will eventually become commonplace.

Outline

- Historical background
- Overview of the Selberg proof
- Overview of the formalization
- Interesting aspects of the formalization
  - Asymptotic reasoning
  - Calculations with reals
  - Casts between natural numbers, integers, and reals
  - Combinatorial reasoning with sums
  - Elementary workarounds
- Heuristic procedures for the reals

Asymptotic reasoning

Define $O(g) = \{f \mid \exists C\ \forall x\ (|f(x)| \leq C|g(x)|)\}$.

Then take "equals" to be "element of" in $f = O(g)$.

The expression makes sense for any function type for which the codomain is an ordered ring.

We used Isabelle's axiomatic type classes to develop the theory in full generality.

Asymptotic reasoning

Define

$$f + g \equiv \lambda x.(f(x) + g(x))$$
$$a +_o B \equiv \{c \mid \exists b \in B\ (c = a + b)\}$$
$$a =_o B \equiv a \in B$$

This gives $f =_o g +_o O(h)$ the intended meaning.

Note that $x^2 + 3x = x^2 + O(x)$ really means

$$(\lambda x.\ x\hat{}2 + 3 * x) =o\ (\lambda x.\ x\hat{}2) +o\ O(\lambda x.\ x)$$

Asymptotic reasoning

Rewrite rules for addition of elements and sets:

set-plus-rearrange	$(a +_o C) + (b +_o D) = (a + b) +_o (C + D)$
set-plus-rearrange2	$a +_o (b +_o C) = (a + b) +_o C$
set-plus-rearrange3	$(a +_o C) + D = a +_o (C + D)$
set-plus-rearrange4	$C + (a +_o D) = a +_o (C + D)$

These put terms in the form $(a + b + \ldots) +_o (C + D + \ldots)$.

Asymptotic reasoning

Some monotonicity and arithmetic rules:

set-plus-intro	$[\|a \in C, b \in D\|] \Rightarrow a + b \in C + D$
set-plus-intro2	$b \in C \Rightarrow a + b \in a + C$
set-plus-mono	$C \subseteq D \Rightarrow a + C \subseteq a + D$
set-plus-mono2	$[\|C \subseteq D, E \subseteq F\|] \Rightarrow C + E \subseteq D + F$
set-plus-mono3	$a \in C \Rightarrow a + D \subseteq C + D$
set-plus-mono4	$a \in C \Rightarrow a + D \subseteq D + C$

Asymptotic reasoning

Some properties of O sets:

bigo-elt-subset	$f \in O(g) \Rightarrow O(f) \subseteq O(g)$
bigo-refl	$f \in O(f)$
bigo-plus-idemp	$O(f) + O(f) = O(f)$
bigo-plus-subset	$O(f + g) \subseteq O(f) + O(g)$
bigo-mult4	$f \in k + oO(h) \Rightarrow g \cdot f \in g \cdot k + oO(g \cdot h)$
bigo-compose1	$f \in O(g) \Rightarrow (\lambda x.\ f(k(x))) \in O(\lambda x.\ g(k(x)))$

Asymptotic reasoning

An annoyance: how do you indicate that $x^3 + 3x^2 + 1 = x^3 + O(x^2)$ for $x \geq 1$?

Options:

1. Define a type of positive reals (or integers).
2. Formalize "$f = O(g)$ on S"
3. Formalize "$f = O(g)$ eventually"
4. Write $(\lambda x.x^3 + 3x^2 + 1) =_o (\lambda x.x^3) +_o O(\lambda x.x^2 + 1)$

We chose the last. This accounts for the endless instances of "+1" and *abs* in our proofs.

The other options have drawbacks, too.

Calculations with real numbers

The very last part of the proof has, by far, the worst length ratio: a difficult 5 page calculation became 89 pages of formal text.

Reason: the need to carry out straightforward calculations by hand, especially involving inequalities.

Isabelle has:

- A term simplifier with ordered rewriting
- Decision procedures of linear and Presburger arithmetic

But lots of easy calculations go just beyond that.

Calculations with real numbers

$$(1 + \frac{\varepsilon}{3(C+3)}) \cdot n < Kx$$

follows from:

$$n \leq (K/2)x$$

$$0 < C$$

$$0 < \varepsilon < 1$$

- Need monotonicity rules for arithmetic operations.
- Need to determine signs.
- Need to remember names like “mult-left-mono.” “add-pos-nonneg,” “order-le-less-trans,” “exp-less-cancel-iff,” “pos-divide-le-eq.”
- Often need to type in long expressions, or cut and paste, or use explicit rules to manipulate terms

Calculations with the real numbers

Sign calculations keep coming back. Consider, for example,

$$1/(1 + st) < 1/(1 + su).$$

These inferences are covered by decision procedures for real closed fields, but

- They are slow.
- Worse: they do not extend to straightforward inferences with monotone functions, trigonometric functions, exponentiation and logarithm, etc.

Consider $x < y \Rightarrow 1/(1 + e^y) < 1/(1 + e^x)$.

Conclusion: we need principled heuristic procedures. (I will come back to this.)

Casting between domains

One can think of θ, ψ, and π as functions from $\mathbb{N}$ to $\mathbb{R}$ or from $\mathbb{R}$ to $\mathbb{R}$.

- Proofs use arithmetic properties of $\mathbb{N}$.
- Ultimately need to cast them to reals.

Recall that μ takes values $\{-1, 0, 1\}$, so we need to deal with integers too.

Casting was an endless source of headaches.

- We had parallel theories of primes and divisibility for ints and nats.
- We had to develop properties of *floor* and *ceiling* functions.
- We had to do annoying manipulations of mixed expressions, e.g. moving $+1$'s in and out of casts, etc.

Casting between domains

When extending a domain (e.g. nats to ints, or ints to reals):

- some operations are extended, like addition and multiplication
- some new operations are mirrored imperfectly in the smaller domain (e.g. $x \dot{-} y$ requires $x \geq y$, $x \mathit{div}\, y$ requires $y|x$).
- some properties depend on the choice of a left inverse, e.g.

$$(n \leq \lfloor x \rfloor) \equiv (\mathit{real}(n) \leq x).$$

The guiding motto should be: anything that is transparent to us should be transparent to a mechanized proof assistant.

Outline

- Historical background
- Overview of the Selberg proof
- Overview of the formalization
- Interesting aspects of the formalization
  - Asymptotic reasoning
  - Calculations with reals
  - Casts between natural numbers, integers, and reals
  - Combinatorial reasoning with sums
  - Elementary workarounds
- Heuristic procedures for the reals

Combinatorial reasoning with sums

Some of our theorems are now in Isabelle's HOL library. For example:

$$\mathit{inj\text{-}on}\ f\ B \Longrightarrow (\sum x{\in}f`B.\ h\ x) = (\sum x{\in}B.\ (h \circ f)(x))$$

"reindexes" a sum.

It is needed, for example, to show

$$\sum_{d|n} h(n) = \sum_{d|n} h(n/d),$$

using $f(d) = n/d$, and

$$\sum_{dd'=c} h(d, d') = \sum_{d|c} h(d, c/d),$$

using $f(d) = \langle d, c/d \rangle$.

Combinatorial reasoning with sums

In the Isabelle formalization, $\sum_{x \in A} f(x)$ is notation for *setsum A f*.

This really only makes sense when A is finite, so finiteness verifications keep popping up in calculations.

(Defining *setsum A f* to be 0 when A is infinite helps.)

According to our motto, there should be better support for finiteness and reindexing.

Elementary workarounds

We relied on the Selberg proof because Isabelle didn't (and still doesn't) have a complex analysis library.

We still don't have a sense of how long it would take to:

- develop a sufficient complex analysis library
- formalize the complex-analytic proof

Of course, the task of finding elementary workarounds is part of the business. It can be oddly enjoyable.

Alas, the need to do this will diminish as formal libraries improve.

Elementary workarounds

Question: how to prove $\ln(1+x) \approx x$ when x is small?

The Isabelle library did not compute the derivative of ln.

It had:

- By definition, $e^x = \sum_{n=0}^{\infty} x^n/n!$
- e^x strictly increasing
- $e^0 = 1$, $e^{x+y} = e^x e^y$
- e^x is surjective on the positive reals
- By definition, $\ln x$ is a left inverse to e^x

Puzzle: show $|\ln(1+x) - x| \leq x^2$ when x is positive and small enough.

Elementary workarounds

Our solution: $x \geq 0$ implies $e^x \geq 1 + x$, so $x \geq \ln(1 + x)$. Replacing x by x^2, we also have $e^{x^2} \geq 1 + x^2$.

On the other hand, the definition of e^x can be used to show

$$e^x \leq 1 + x + x^2$$

when $0 \leq x \leq 1/2$. From these we get

$$e^{x-x^2} = e^x / e^{x^2} \leq (1 + x + x^2)/(1 + x^2) \leq 1 + x.$$

Taking logarithms of both sides, we have

$$x - x^2 \leq \ln(1 + x) \leq x$$

when $0 \leq x \leq 1/2$, as required.

Elementary workarounds

Another puzzle: show

$$\sum_{n \le x} 1/n = \ln x + O(1)$$

without integration. When x is positive, write

$$\begin{aligned}
\ln x &= \sum_{n \le x-1} (\ln(n+1) - \ln n) \\
&= \sum_{n \le x-1} \ln(1 + 1/n) \\
&= \sum_{n \le x-1} 1/n + O(\sum_{n \le x} 1/n^2) \\
&= \sum_{n \le x} 1/n + O(1).
\end{aligned}$$

Outline

- Historical background
- Overview of the Selberg proof
- Overview of the formalization
- Interesting aspects of the formalization
  - Asymptotic reasoning
  - Calculations with reals
  - Casts between natural numbers, integers, and reals
  - Combinatorial reasoning with sums
  - Elementary workarounds
- Heuristic procedures for the reals

Heuristic procedures for the reals

Remember the example: verify

$$(1 + \frac{\varepsilon}{3(C+3)}) \cdot n < Kx$$

using the following hypotheses:

$$n \leq (K/2)x$$
$$0 < C$$
$$0 < \varepsilon < 1$$

Idea: work backwards, applying obvious monotonicity rules.

Heuristic procedures for the reals

Problems:

1. Case splits: e.g. $st > 0 \equiv (s > 0 \wedge t > 0) \vee (s < 0 \wedge t < 0)$.
2. Nondeterminism: e.g. many ways to show $s + t < u + v + w$.

Observations:

1. "Straightforward" inferences usually don't need case splits.
2. In practice, Fourier-Motzkin is efficient for linear inequalities.
3. Modulo cases over signs, the same thing works for the multiplicative fragment of the reals.

Heuristic procedures for the reals

Let T_1 be the theory of $\langle \mathbb{R}, 0, +, < \rangle$. T_1 is decidable.

Let T_2 be the theory of $\langle \mathbb{R}, 1, \times, < \rangle$. T_2 is decidable.

Let $T = T_1 \cup T_2$. By Nelson-Oppen methods, the universal fragment of T is decidable.

Problem: T is too weak; it doesn't prove $2 \times 2 = 4$.

Heuristic procedures for the reals

A better version: let $f_a(x) = ax$ for rational constants a.

Let $T_1[\mathbb{Q}]$ be the theory of $\langle \mathbb{R}, 0, 1, +, -, <, \ldots, f_a, \ldots \rangle$.

Let $T_2[\mathbb{Q}]$ be the theory of $\langle \mathbb{R}, 0, 1, \times, \div, \sqrt[n]{\cdot}, <, \ldots, f_a, \ldots \rangle$.

Let $T[\mathbb{Q}] = T_1[\mathbb{Q}] \cup T_2[\mathbb{Q}]$.

Both of these are decidable, but Nelson-Oppen methods fail when there is a nontrivial overlap.

The situation here is much more complex!

Heuristic procedures for the reals

This is joint work with Harvey Friedman.

Here are some things we (think we) know:

- $T[\mathbb{Q}]$ has good normal forms.
- Valid equations are independent of the ordering.
- $T[\mathbb{Q}]$ is undecidable.
- In fact, the $\forall\forall\forall\exists \ldots \exists$ fragment is complete r.e.
- Assuming that the solvability of Diophantine equations in the rationals is undecidable, then so is the existential fragment of $T[\mathbb{Q}]$.
- The universal fragment of $T[\mathbb{Q}]$ *is* decidable.

We have similar results, for example, with the real algebraic numbers A in place of $\mathbb{Q}$.

Heuristic procedures for the reals

Our decidability results are not practical. But the proofs provide ideas and guidelines.

General strategy for amalgamation:

- Maintain a database of facts in the common language.
- Iteratively use each of T_1 and T_2 to add new facts.

Issues:

- Heuristically, how to decide *which* facts to focus on?
- When to split on cases?
- How to look for disjunctions?
- How to incorporate distributivity?
- How to amalgamate other local decision or heuristic procedures?

Conclusions

Formally verified mathematics is becoming increasingly important:

- Proofs are getting very complex.
- Proofs rely on extensive computations.

Fortunately, we are entering "the golden age of metamathematics" (Shankar).

Continued progress will require

- thoughtful reflection
- good theory
- solid engineering

This makes the field an auspicious combination of theory and practice.