Ramanujan and Erdös – Two Proofs of Bertrand's Postulate

March 23, 2012

Joseph Bertrand conjectured the following postulate in 1845 after testing it for numbers till three million:[1]

For every n, there is a prime betwee n and $2n$.

The first proof was given by Chebyshev, followed by an elementary proof by Ramanujan in 1919, which instead was follows by an elementary proof by Erdös (in his first paper at the age of 19!). We will see the latter two proofs. Both have their strengths and drawbacks. Erdös proof uses nothing more than some astute observations on the factorization of the largest binomial coefficient, and derives a contradiction. Ramanujan's proof, on the other hand, not only shows the existence of primes between n and $2n$, but says something more: the number of primes between n and $2n$ increase, as n increases.

Before we see these proofs, we derive some upper and lower bounds on the largest binomial coefficient. These bounds are crucial in both the proofs. As a warmup, we start with estimating general binomial coefficients.

1 Estimates on Binomial Coefficients

Let's consider the binomial coefficient $\binom{n}{k}$. We know from the Binomial theorem that: $\sum_{k=0}^{n}\binom{n}{k} = 2^n$. So clearly, $\binom{n}{k} \le 2^n$. But that's not a very tight bound. Our first claim improves upon that.

Claim 1:

$$\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \frac{n^k}{2^{k-1}}.$$

The proof follows from the definition

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!}.$$

The RHS of the claim follows from the fact that all there are k terms in the numerator, each not exceeding n, and the denominator $k! \ge 2^{k-1}$. The LHS of the claim follows from the observation that for $0 \le i < k$, $(n-i)/(k-i) \ge n/k$.

The next claim improves the upper bound.

Claim 2:

$$\binom{n}{k} \le \left(\frac{ne}{k}\right)^k.$$

We will show a stronger claim: $\sum_{j=0}^{k}\binom{n}{j} < (ne/k)^k$. However, this bound is not always better than the naive bound of 2^n, since for $k = n$, we get an upper bound of $(2e)^n$; even for $k = n/2$, the bound is $(2e)^{n/2}$ and is still not better than 2^{n-1}; the bound is better than $2^{\epsilon n}$, where $0 < \epsilon < 1$, if $k = O(n/(\epsilon \ln n))$. To prove our stronger claim, we start with the Binomial theorem and observe that for $x \ge 0$.

$$\sum_{j=0}^{k}\binom{n}{j}x^j \le (1+x)^n.$$

[1]It is called a postulate rather than "conjecture" because he took it as a working tool in his study of a problem in group theory. This must have seemed entirely safe, considering the actual density of primes in the tables. There is not merely one prime between 500,000 and 1,000,000, say, there are 36,960 of them! – W J LeVeque, Fundamentals of Number Theory.

Dividing both sides by x^k, we obtain

$$\sum_{j=0}^{k} \binom{n}{j} \frac{1}{x^{k-j}} \leq \frac{(1+x)^n}{x^k}.$$

Now, if for all $x \leq 1$ the summation on the LHS is greater than $\sum_{j=0}^{k} \binom{n}{j}$. For what value of $x < 1$, does the term on the RHS obtain its minimum. Using simple calculus we see that the term attains its minimum at the point $x = k/(n-k)$. Since the value attained by the term at $x = k/(n-k)$ is smaller that the value attained at $x = k/n$, we obtain

$$\sum_{j=0}^{k} \binom{n}{j} \leq \left(1 + \frac{k}{n}\right)^n \left(\frac{n}{k}\right)^k.$$

Since the function $e^x > (1 + x/n)^n$, we get the desired upper bound

$$\sum_{j=0}^{k} \binom{n}{j} \leq \left(\frac{ne}{k}\right)^k.$$

We are interested in deriving a good estimate on $B_n := \binom{2n}{n}$. As mentioned earlier, from the second claim we obtain that $B_n \leq (2e)^{n/2}$, which is not that good. For a lower bound, we can proceed as follows: since $\sum_{i=0}^{2n} \binom{2n}{i} = 2^{2n}$, the average of the $2n$ numbers $\binom{2n}{0} + \binom{2n}{1}, \binom{2n}{2}, \ldots, \binom{2n}{2n}$ is $2^{2n}/2n$; but clearly the maximum amongst these $2n$ numbers, namely B_n, is larger than the average, i.e.

$$B_n \geq \frac{2^{2n}}{2n}. \tag{1}$$

The following claim improves upon both the upper and lower bounds.

Claim 3:

$$\frac{2^{2n}}{2\sqrt{n}} < \binom{2n}{n} < \frac{2^{2n}}{\sqrt{2n}}. \tag{2}$$

Let

$$A := \binom{2n}{n} \frac{1}{2^{2n}}.$$

Then our claim follows if we show that $1/(2\sqrt{n}) \leq A \leq 1/\sqrt{2n}$, or equivalently

$$\frac{1}{4n} < A^2 < \frac{1}{2n}.$$

So let's consider the quantity

$$\begin{aligned}
A^2 = \frac{(2n!)^2}{(n!)^4} 2^{-4n} &= \left(\frac{2n(2n-1)\ldots 2}{(n(n-1)\ldots 2)^2}\right)^2 \frac{1}{2^{4n}} \\
&= \left(\frac{2n(2n-1)\ldots 2}{(2^n \cdot n(n-1)\ldots 2)^2}\right)^2 \\
&= \left(\frac{2n(2n-1)\ldots 2}{(2n)^2(2(n-1))^2\ldots(2\cdot 2)^2 2^2}\right)^2 \\
&= \left(\frac{(2n-1)(2n-3)\ldots 5\cdot 3}{(2n)(2(n-1))\ldots(2\cdot 2)\cdot 2}\right)^2 \\
&= \frac{(2n-1)}{(2n)^2} \frac{(2n-1)(2n-3)}{(2(n-1))^2} \ldots \frac{5\cdot 3}{4^2} \frac{3\cdot 1}{2^2}.
\end{aligned}$$

Thus

$$\begin{aligned}
2nA^2 &= \frac{(2n)(2n-1)}{(2n)^2} \frac{(2n-1)(2n-3)}{(2(n-1))^2} \ldots \frac{5\cdot 3}{4^2} \frac{3\cdot 1}{2^2} \\
&= \left(1 - \frac{1}{(2n)^2}\right)\left(1 - \frac{1}{(2(n-1))^2}\right) \ldots \left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{2^2}\right) \\
&< 1
\end{aligned}$$

which gives us the desired upper bound on A^2. To derive the lower bound we observe that

$$\begin{aligned}\frac{1}{A^2} &= \left(\frac{(2n)(2(n-1))\dots(2\cdot 2)\cdot 2}{(2n-1)(2n-3)\dots 5\cdot 3}\right)^2 \\ &= 2n\cdot\frac{(2n)(2(n-1))}{(2n-1)^2}\frac{(2(n-1))(2(n-2))}{(2n-3)^2}\dots\frac{6\cdot 4}{5^2}\frac{4\cdot 2}{3^2}\cdot 2 \\ &= 4n\left(1-\frac{1}{(2n-1)^2}\right)\left(1-\frac{1}{(2n-3)^2}\right)\dots\left(1-\frac{1}{5^2}\right)\left(1-\frac{1}{3^2}\right) \\ &< 4n,\end{aligned}$$

which gives us the desired lower bound on A^2.

In the following sections, p is always taken to be a prime; consequently, $\sum_p$ and $\prod_p$ represent sums and products over primes.

2 Erdös's Proof

The proof is based upon the prime factorisation of the binomial coefficient $B_n := \binom{2n}{n}$. We first observe that any prime that divides B_n is smaller than $2n$, since in the numerator $2n!$ the largest number is at most $2n$. Let $\mathtt{ord}_p(m)$ denote the largest power of p that divides m; it's straightforward to verify that $\mathtt{ord}_p(a/b) = \mathtt{ord}_p(a) - \mathtt{ord}_p(b)$ and $\mathtt{ord}_p(ab) = \mathtt{ord}_p(a) + \mathtt{ord}_p(b)$. Given this notation, we can write the prime factorisation of B_n as follows:

$$B_n = \prod_{p\le\sqrt{2n}} p^{\mathtt{ord}_p(B_n)} \prod_{\sqrt{2n}<p\le 2n/3} p^{\mathtt{ord}_p(B_n)} \prod_{2n/3<p\le n} p^{\mathtt{ord}_p(B_n)} \prod_{n<p\le 2n} p^{\mathtt{ord}_p(B_n)}.$$

Erdös made the following crucial observation: there cannot be any prime p, $2n/3 < p \le n$, that divides B_n; because $\mathtt{ord}_p(2n!) = 2$ (only p and $2p$ appears in $2n!$, since $2n < 3p$) and $\mathtt{ord}_p(n) = 1$ (only p appears in $n!$, as $n < 1.5p$); thus, $\mathtt{ord}_p(B_n) = \mathtt{ord}_p(2n!) - 2\mathtt{ord}_p(n!) = 0$. The proof of the postulate is by contradiction. If the postulate was false then the factorisation of B_n is equal to

$$B_n = \prod_{p\le\sqrt{2n}} p^{\mathtt{ord}_p(B_n)} \prod_{\sqrt{2n}<p\le 2n/3} p^{\mathtt{ord}_p(B_n)}. \tag{3}$$

Since all the prime factors are smaller than $2n/3$, the equation above suggests that B_n cannot be very large. In fact, we will derive an upper bound on B_n that is smaller than a lower bound, hence giving us a contradiction.

We will need the following observation.

THEOREM 1 (Legendre's Theorem). *The order of p in $n!$ is*

$$\mathtt{ord}_p(n!) = \sum_{k\ge 1}\left\lfloor\frac{n}{p^k}\right\rfloor.$$

Proof. We count the contribution of p to $\mathtt{ord}_p(n!)$, namely $p, 2p, 3p\dots$. The number of such terms are $\lfloor n/p\rfloor$. Similarly, the contribution of p^2 to the order is $\lfloor n/p^2\rfloor$; note that we are not overcounting the order, since we count one each for the contribution of $p, p^2, p^3\dots$. **Q.E.D.**

From the theorem above, we get the following:

$$\begin{aligned}\mathtt{ord}_p(B_n) &= \mathtt{ord}_p(2n!) - 2\mathtt{ord}_p(n!) \\ &= \textstyle\sum_{k=1}^{\lfloor\log_p(2p)\rfloor}\left(\left\lfloor\frac{2n}{p^k}\right\rfloor - 2\left\lfloor\frac{n}{p^k}\right\rfloor\right).\end{aligned}$$

Consider the term on the RHS in the brackets. We claim that it's an integer not greater than one, since

$$\left\lfloor\frac{2n}{p^k}\right\rfloor - 2\left\lfloor\frac{n}{p^k}\right\rfloor \le \frac{2n}{p^k} - 2\left\lfloor\frac{n}{p^k}\right\rfloor < \frac{2n}{p^k} - 2\left(\frac{n}{p^k}-1\right) = 2.$$

Thus, for all primes p, $\mathtt{ord}_p(B_n) \leq \log_p(2p)$, or in other words

$$p^{\mathtt{ord}_p(B_n)} \leq 2n. \tag{4}$$

From (4) we get that

$$B_n \leq (2n)^{\sqrt{2n}} \prod_{\sqrt{2n}<p\leq 2n/3} p^{\mathtt{ord}_p(B_n)}.$$

From (4) we also get that for p in the range $(\sqrt{2n}, 2n/3]$, $\mathtt{ord}_p(B_n) \leq 1$. Thus

$$B_n \leq (2n)^{\sqrt{2n}} \prod_{\sqrt{2n}<p\leq 2n/3} p.$$

From the equation above and (2), we obtain

$$\frac{2^{2n}}{2n} < (2n)^{\sqrt{2n}} \prod_{\sqrt{2n}<p\leq 2n/3} p. \tag{5}$$

Suppose we substitute the upperbound $(2n/3)^{2n/3}$ on the product on the RHS. Then we don't get a contradiction, since asymptotically the RHS dominates the LHS; though, note that the LHS dominates the term $(2n)^{\sqrt{2n}}$ on the RHS. We thus need a tighter upperbound on the product; the lower bound suggests that we need an upper bound of the form c^n. The following lemma gives us the desired upperbound.

LEMMA 2. *The product of all primes not exceeding m is at most 4^m.*

Proof. We prove it by induction on m; the base case $m = 2$ is clear. Suppose the argument holds for all numbers less than m.

If m is even then we know that

$$\prod_{p\leq m} p = \prod_{p\leq m-1} p \leq 4^{m-1}.$$

If m is odd the same argument does not work, as m may be a prime. Let $m = 2k + 1$. Then

$$\prod_{p\leq m} p = \prod_{p\leq k+1} p \prod_{k+1<p\leq 2k+1} p \leq 4^{k+1} \prod_{k+1<p\leq 2k+1} p,$$

where the inequality follows from the inductive hypothesis. The product on the RHS divides the binomial $\binom{2k+1}{k}$, since none of the primes occur in the denominator. Thus

$$\prod_{p\leq m} p \leq 4^{k+1}\binom{2k+1}{k}.$$

We know that $\binom{2k+1}{k} + \binom{2k+1}{k+1} \leq 2^{2k+1}$; since the two binomial coefficients are equal, we get $\binom{2k+1}{k} \leq 2^{2k}$. Thus we have the desired upperbound

$$\prod_{p\leq m} p \leq 4^{k+1}4^k = 4^{2k+1}.$$

Q.E.D.

Using the theorem above in (5), we get that if the postulate is not true then n must satisfy

$$\frac{4^n}{2n} \leq (2n)^{\sqrt{2n}} 4^{2n/3},$$

or equivalently

$$4^{n/3} \leq (2n)^{\sqrt{2n}+1},$$

which implies that $n < 500$; the above inequality cannot hold asymptotically, since if we take logarithms on both sides we get $n = O(\sqrt{n}\log n)$. Thus our assumption leads us to the claim that the postulate is not true when $n < 500$. However, this is not true because we can construct a list of primes less than 500 such that the $(i+1)$th prime is smaller than twice the ith prime. Thus our assumption is not true, and hence the postulate must hold for all n.

3 Ramanujan's Proof

Instead of considering the product of primes, Ramnujan considered the so called Chebyshev function $\theta(x) := \sum_{p \le x} \ln p$, where x is a positive real number. We say, Bertrand's postulate is true for x if it is true for $n := \lfloor x \rfloor$, since if there is a prime between n and $2n$ then there is a prime between x and $2x$; note it's not an iff statement. To prove Bertrand's postulate, Ramanujan demonstrates that for all $x \ge 1$, $\theta(2x) - \theta(x) \ge 1$. This he obtained by deriving good upper and lower bounds on $\theta(x)$. Though we already have the upper bound $\theta(x) < x \ln 2$ from Lemma 2, there is no evident way of deriving a corresponding lower bound based upon similar same arguments. Ramanujan's approach to getting the bounds is slightly indirect: he introduced an intermediate function $\Psi(x)$, which we will see later, such that $\theta(x)$ has an upper and lower bound in terms of $\Psi(x)$; then he derives an upper bound on $\Psi(x)$, in fact the upper bound is $x \ln 2$, and uses it to get the lower bound for $\theta(x)$. To derive bounds on $\Psi(x)$, he derives a relation of $\Psi(x)$ with the logarithm of the familiar binomial coefficient B_n. This section is based upon [1].

We begin with a reformulation of Legendre's theorem, Theorem 1. If $m := \left\lfloor \frac{n}{p^k} \right\rfloor$ then

$$mp^k \le n < (m+1)p^k.$$

From this observation we see that

$$m = \sum_{i \ge 1} [ip^k \le n]$$

where we use Iverson's convention. Using this equality, Legendre's theorem can be restated as

$$\mathtt{ord}_p(n!) = \sum_{k \ge 1} \sum_{i \ge 1} [ip^k \le n].$$

Therefore,

$$\begin{aligned}
\ln n! &= \sum_{p \le n} \mathtt{ord}_p(n!) \ln p \\
&= \sum_{p \le n} \sum_{k \ge 1} \sum_{i \ge 1} [ip^k \le n] \ln p \\
&= \sum_{k \ge 1} \sum_{i \ge 1} \sum_{p \le n} [ip^k \le n] \ln p \\
&= \sum_{k \ge 1} \sum_{i \ge 1} \sum_{p : p \le (n/i)^{1/k}} \ln p.
\end{aligned}$$

Using the Chebyshev function, and swapping the order of summation over k and i, the above result can be expressed as

$$\ln n! = \sum_{i \ge 1} \sum_{k \ge 1} \theta \left(\left(\frac{n}{i} \right)^{1/k} \right).$$

Further define

$$\Psi(x) := \sum_{k \ge 1} \theta(x^{1/k}) = \theta(x) + \theta(x^{1/2}) + \theta(x^{1/3}) + \cdots . \tag{6}$$

Then

$$\ln n! = \sum_{i \ge 1} \Psi(n/i),$$

and in general for any x,

$$\ln \lfloor x \rfloor ! = \sum_{i \ge 1} \Psi(\lfloor x \rfloor / i) = \sum_{i \ge 1} \Psi(x/i) = \Psi(x) + \Psi(x/2) + \Psi(x/3) + \cdots . \tag{7}$$

From (6) it follows that

$$\Psi(\sqrt{x}) = \sum_{k \ge 1} \theta(x^{1/2k}),$$

and hence

$$\Psi(x) - 2\Psi(\sqrt{x}) = \theta(x) - \theta(x^{1/2}) + \theta(x^{1/3}) - \theta(x^{1/4}) + \cdots . \tag{8}$$

Since $\theta(x)$ is an increasing function it immediately follows that

$$\Psi(x) - 2\Psi(\sqrt{x}) < \theta(x) < \Psi(x) \tag{9}$$

where the upper bound is from (6). In order to bound $\Psi(x)$, we relate it to the familiar binomial coefficient. From (7), it follows that

$$\ln \lfloor x/2 \rfloor ! = \sum_{i \geq 1} \Psi(x/2i)$$

and hence

$$\ln \binom{\lfloor x \rfloor}{\lfloor x/2 \rfloor} = \ln \lfloor x \rfloor ! - 2 \ln \lfloor x/2 \rfloor ! = \Psi(x) - \Psi(x/2) + \Psi(x/3) - \Psi(x/4) + \cdots .$$

Since $\Psi(x)$ is an increasing function, we have

$$\Psi(x) - \Psi(x/2) < \ln \lfloor x \rfloor ! - 2 \ln \lfloor x/2 \rfloor ! < \Psi(x) - \Psi(x/2) + \Psi(x/3).$$

Using the trivial upper bound $\binom{\lfloor x \rfloor}{\lfloor x/2 \rfloor} \leq 2^x$, we obtain

$$\Psi(x) - \Psi(x/2) < x \ln 2$$

and from the lower bound in (2) we get

$$x \ln 2 - \ln \sqrt{2x} < \Psi(x) - \Psi(x/2) + \Psi(x/3). \tag{10}$$

Let's consider the first inequality. Substituting $x/2, x/4, \ldots,$ on the LHS and adding all the terms we obtain

$$\Psi(x) < x \ln 2 \ (1 + 1/2 + 1/4 + \cdots) = 2x \ln 2. \tag{11}$$

Furthermore, from (9), we know

$$\theta(x) < \Psi(x) < 2x \ln 2,$$

giving us a good upper bound on $\theta(x)$ (which is not surprising, as mentioned earlier). To derive a similar lower bound on $\theta(x) - \theta(x/2)$, we again apply the upper and lower bounds in (9) to obtain

$$\theta(x) - \theta(x/2) > \theta(x) - \Psi(x/2) > \Psi(x) - 2\Psi(\sqrt{x}) - \Psi(x/2).$$

Inequality (10) further implies that

$$\theta(x) - \theta(x/2) > x \ln 2 - \ln \sqrt{2x} - \Psi(x/3) - 2\Psi(\sqrt{x}).$$

Plugging the upper bound on $\Psi(x)$ from (11) in the RHS we get

$$\theta(x) - \theta(x/2) > x \ln 2 - \ln \sqrt{2x} - \frac{2}{3} x \ln 2 - 2\sqrt{x} \ln 2 = \frac{x}{3} \ln 2 - \ln \sqrt{2x} - 2\sqrt{x} \ln 2.$$

For $x \geq 80$ the RHS is greater tha one; in fact it is an increasing unbounded function of x; for $x < 80$ we can verify that the postulate is true. Thus the postulate holds for all positive real numbers greater than one.

What next? Is Legendre's claim true: for all $n \geq 1$, there is a prime between n^2 and $(n+1)^2$?

References

[1] B. Sury. How Far Apart are Primes? Bertrand's Postulate. *Resonance*, pages 77–87, 2002.