Bertrand's postulate

Chapter 2

We have seen that the sequence of prime numbers $2, 3, 5, 7, \ldots$ is infinite. To see that the size of its gaps is not bounded, let $N := 2 \cdot 3 \cdot 5 \cdots p$ denote the product of all prime numbers that are smaller than $k + 2$, and note that none of the k numbers

$$N + 2, N + 3, N + 4, \ldots, N + k, N + (k + 1)$$

is prime, since for $2 \leq i \leq k + 1$ we know that i has a prime factor that is smaller than $k + 2$, and this factor also divides N, and hence also $N + i$. With this recipe, we find, for example, for $k = 10$ that none of the ten numbers

$$2312, 2313, 2314, \ldots, 2321$$

is prime.

But there are also upper bounds for the gaps in the sequence of prime numbers. A famous bound states that "the gap to the next prime cannot be larger than the number we start our search at." This is known as Bertrand's postulate, since it was conjectured and verified empirically for $n < 3\,000\,000$ by Joseph Bertrand. It was first proved for all n by Pafnuty Chebyshev in 1850. A much simpler proof was given by the Indian genius Ramanujan. Our Book Proof is by Paul Erdős: it is taken from Erdős' first published paper, which appeared in 1932, when Erdős was 19.

Joseph Bertrand

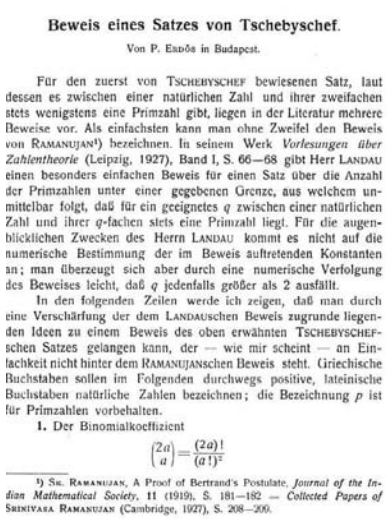
Bertrand's postulate.
For every $n \geq 1$, there is some prime number p with $n < p \leq 2n$.

■ **Proof.** We will estimate the size of the binomial coefficient $\binom{2n}{n}$ carefully enough to see that if it didn't have any prime factors in the range $n < p \leq 2n$, then it would be "too small." Our argument is in five steps.

(1) We first prove Bertrand's postulate for $n < 4000$. For this one does not need to check 4000 cases: it suffices (this is "Landau's trick") to check that

$$2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 631, 1259, 2503, 4001$$

is a sequence of prime numbers, where each is smaller than twice the previous one. Hence every interval $\{y : n < y \leq 2n\}$, with $n \leq 4000$, contains one of these 14 primes.



Beweis eines Satzes von Tschebyschef.

Von P. Erdös in Budapest.

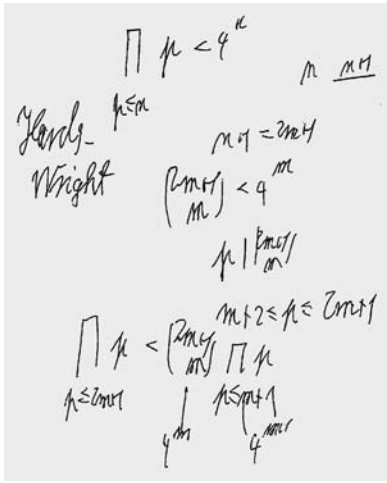
Für den zuerst von Tschebyschef bewiesenen Satz, laut dessen es zwischen einer natürlichen Zahl und ihrer zweifachen stets wenigstens eine Primzahl gibt, liegen in der Literatur mehrere Beweise vor. Als einfachsten kann man ohne Zweifel den Beweis von Ramanujan[1]) bezeichnen. In seinem Werk *Vorlesungen über Zahlentheorie* (Leipzig, 1927), Band I, S. 66—68 gibt Herr Landau einen besonders einfachen Beweis für einen Satz über die Anzahl der Primzahlen unter einer gegebenen Grenze, aus welchem unmittelbar folgt, daß für ein geeignetes q zwischen einer natürlichen Zahl und ihrer q-fachen stets eine Primzahl liegt. Für die augenblicklichen Zwecken des Herrn Landau kommt es nicht auf die numerische Bestimmung der im Beweis auftretenden Konstanten an; man überzeugt sich aber durch eine numerische Verfolgung des Beweises leicht, daß q jedenfalls größer als 2 ausfällt.

In den folgenden Zeilen werde ich zeigen, daß man durch eine Verschärfung der dem Landauschen Beweis zugrunde liegenden Ideen zu einem Beweis des oben erwähnten Tschebyschefschen Satzes gelangen kann, der — wie mir scheint — an Einfachkeit nicht hinter dem Ramanujanschen Beweis steht. Griechische Buchstaben sollen im Folgenden durchwegs positive, lateinische Buchstaben natürliche Zahlen bezeichnen; die Bezeichnung p ist für Primzahlen vorbehalten.

1. Der Binomialkoeffizient

$$\binom{2a}{a} = \frac{(2a)!}{(a!)^2}$$

[1]) Sr. Ramanujan, A Proof of Bertrand's Postulate, *Journal of the Indian Mathematical Society*, 11 (1919), S. 181—182 — *Collected Papers of* Srinivasa Ramanujan (Cambridge, 1927), S. 208—209.



(2) Next we prove that

$$\prod_{p\le x} p \;\le\; 4^{x-1} \qquad \text{for all real } x \ge 2, \tag{1}$$

where our notation — here and in the following — is meant to imply that the product is taken over all *prime* numbers $p \le x$. The proof that we present for this fact uses induction on the number of these primes. It is not from Erdős' original paper, but it is also due to Erdős (see the margin), and it is a true Book Proof. First we note that if q is the largest prime with $q \le x$, then

$$\prod_{p\le x} p \;=\; \prod_{p\le q} p \qquad \text{and} \qquad 4^{q-1} \;\le\; 4^{x-1}.$$

Thus it suffices to check (1) for the case where $x = q$ is a prime number. For $q = 2$ we get "$2 \le 4$," so we proceed to consider odd primes $q = 2m+1$. (Here we may assume, by induction, that (1) is valid for all integers x in the set $\{2, 3, \dots, 2m\}$.) For $q = 2m+1$ we split the product and compute

$$\prod_{p\le 2m+1} p \;=\; \prod_{p\le m+1} p \cdot \prod_{m+1<p\le 2m+1} p \;\le\; 4^m \binom{2m+1}{m} \;\le\; 4^m 2^{2m} \;=\; 4^{2m}.$$

All the pieces of this "one-line computation" are easy to see. In fact,

$$\prod_{p\le m+1} p \;\le\; 4^m$$

holds by induction. The inequality

$$\prod_{m+1<p\le 2m+1} p \;\le\; \binom{2m+1}{m}$$

follows from the observation that $\binom{2m+1}{m} = \frac{(2m+1)!}{m!(m+1)!}$ is an integer, where the primes that we consider all are factors of the numerator $(2m+1)!$, but not of the denominator $m!(m+1)!$. Finally

$$\binom{2m+1}{m} \;\le\; 2^{2m}$$

holds since

$$\binom{2m+1}{m} \text{ and } \binom{2m+1}{m+1}$$

are two (equal!) summands that appear in

$$\sum_{k=0}^{2m+1} \binom{2m+1}{k} = 2^{2m+1}.$$

(3) From Legendre's theorem (see the box) we get that $\binom{2n}{n} = \frac{(2n)!}{n!n!}$ contains the prime factor p exactly

$$\sum_{k\ge 1} \left(\left\lfloor \frac{2n}{p^k} \right\rfloor - 2 \left\lfloor \frac{n}{p^k} \right\rfloor \right)$$

Legendre's theorem

The number $n!$ contains the prime factor p exactly

$$\sum_{k\ge 1} \left\lfloor \frac{n}{p^k} \right\rfloor$$

times.

■ **Proof.** Exactly $\lfloor \frac{n}{p} \rfloor$ of the factors of $n! = 1 \cdot 2 \cdot 3 \cdots n$ are divisible by p, which accounts for $\lfloor \frac{n}{p} \rfloor$ p-factors. Next, $\lfloor \frac{n}{p^2} \rfloor$ of the factors of $n!$ are even divisible by p^2, which accounts for the next $\lfloor \frac{n}{p^2} \rfloor$ prime factors p of $n!$, etc. □

times. Here each summand is at most 1, since it satisfies

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

and it is an integer. Furthermore the summands vanish whenever $p^k > 2n$. Thus $\binom{2n}{n}$ contains p exactly

$$\sum_{k\geq 1}\left(\left\lfloor \frac{2n}{p^k} \right\rfloor - 2\left\lfloor \frac{n}{p^k} \right\rfloor\right) \;\leq\; \max\{r : p^r \leq 2n\}$$

times. Hence the largest power of p that divides $\binom{2n}{n}$ is not larger than $2n$. In particular, primes $p > \sqrt{2n}$ appear at most once in $\binom{2n}{n}$.

Furthermore — and this, according to Erdős, is the key fact for his proof — primes p that satisfy $\frac{2}{3}n < p \leq n$ do not divide $\binom{2n}{n}$ at all! Indeed, $3p > 2n$ implies (for $n \geq 3$, and hence $p \geq 3$) that p and $2p$ are the only multiples of p that appear as factors in the numerator of $\frac{(2n)!}{n!n!}$, while we get two p-factors in the denominator.

Examples such as
$\binom{26}{13} = 2^3 \cdot 5^2 \cdot 7 \cdot 17 \cdot 19 \cdot 23$
$\binom{28}{14} = 2^3 \cdot 3^3 \cdot 5^2 \cdot 17 \cdot 19 \cdot 23$
$\binom{30}{15} = 2^4 \cdot 3^2 \cdot 5 \cdot 17 \cdot 19 \cdot 23 \cdot 29$
illustrate that "very small" prime factors $p < \sqrt{2n}$ can appear as higher powers in $\binom{2n}{n}$, "small" primes with $\sqrt{2n} < p \leq \frac{2}{3}n$ appear at most once, while factors in the gap with $\frac{2}{3}n < p \leq n$ don't appear at all.

(4) Now we are ready to estimate $\binom{2n}{n}$. For $n \geq 3$, using an estimate from page 12 for the lower bound, we get

$$\frac{4^n}{2n} \;\leq\; \binom{2n}{n} \;\leq\; \prod_{p\leq\sqrt{2n}} 2n \;\cdot \prod_{\sqrt{2n}<p\leq\frac{2}{3}n} p \;\cdot \prod_{n<p\leq 2n} p$$

and thus, since there are not more than $\sqrt{2n}$ primes $p \leq \sqrt{2n}$,

$$4^n \;\leq\; (2n)^{1+\sqrt{2n}} \cdot \prod_{\sqrt{2n}<p\leq\frac{2}{3}n} p \;\cdot \prod_{n<p\leq 2n} p \qquad \text{for} \quad n \geq 3. \tag{2}$$

(5) Assume now that there is no prime p with $n < p \leq 2n$, so the second product in (2) is 1. Substituting (1) into (2) we get

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

or

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

which is false for n large enough! In fact, using $a + 1 < 2^a$ (which holds for all $a \geq 2$, by induction) we get

$$2n = \left(\sqrt[6]{2n}\right)^6 < \left(\left\lfloor \sqrt[6]{2n} \right\rfloor + 1\right)^6 < 2^{6\left\lfloor \sqrt[6]{2n} \right\rfloor} \leq 2^{6\sqrt[6]{2n}}, \tag{4}$$

and thus for $n \geq 50$ (and hence $18 < 2\sqrt{2n}$) we obtain from (3) and (4)

$$2^{2n} \leq (2n)^{3\left(1+\sqrt{2n}\right)} < 2^{\sqrt[6]{2n}\left(18+18\sqrt{2n}\right)} < 2^{20\sqrt[6]{2n}\sqrt{2n}} = 2^{20(2n)^{2/3}}.$$

This implies $(2n)^{1/3} < 20$, and thus $n < 4000$. □

One can extract even more from this type of estimates: From (2) one can derive with the same methods that

$$\prod_{n<p\le 2n} p \ge 2^{\frac{1}{30}n} \quad \text{for} \quad n \ge 4000,$$

and thus that there are at least

$$\log_{2n}\left(2^{\frac{1}{30}n}\right) \;=\; \frac{1}{30}\,\frac{n}{\log_2 n+1}$$

primes in the range between n and $2n$.

This is not that bad an estimate: the "true" number of primes in this range is roughly $n/\log n$. This follows from the "prime number theorem," which says that the limit

$$\lim_{n\to\infty} \frac{\#\{p\le n : p \text{ is prime}\}}{n/\log n}$$

exists, and equals 1. This famous result was first proved by Hadamard and de la Vallée-Poussin in 1896; Selberg and Erdős found an elementary proof (without complex analysis tools, but still long and involved) in 1948.

On the prime number theorem itself the final word, it seems, is still not in: for example a proof of the Riemann hypothesis (see page 49), one of the major unsolved open problems in mathematics, would also give a substantial improvement for the estimates of the prime number theorem. But also for Bertrand's postulate, one could expect dramatic improvements. In fact, the following is a famous unsolved problem:

Is there always a prime between n^2 and $(n+1)^2$?

For additional information see [3, p. 19] and [4, pp. 248, 257].

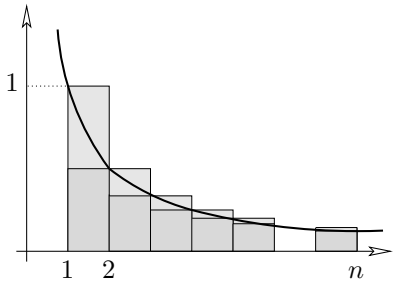
Appendix: Some estimates

Estimating via integrals

There is a very simple-but-effective method of estimating sums by integrals (as already encountered on page 4). For estimating the *harmonic numbers*

$$H_n \;=\; \sum_{k=1}^{n} \frac{1}{k}$$

we draw the figure in the margin and derive from it



$$H_n - 1 \;=\; \sum_{k=2}^{n} \frac{1}{k} \;<\; \int_1^n \frac{1}{t}\,dt \;=\; \log n$$

by comparing the area below the graph of $f(t)=\frac{1}{t}$ $(1\le t\le n)$ with the area of the dark shaded rectangles, and

$$H_n - \frac{1}{n} \;=\; \sum_{k=1}^{n-1} \frac{1}{k} \;>\; \int_1^n \frac{1}{t}\,dt \;=\; \log n$$

by comparing with the area of the large rectangles (including the lightly shaded parts). Taken together, this yields

$$\log n + \frac{1}{n} \;<\; H_n \;<\; \log n \;+\; 1.$$

In particular, $\lim_{n\to\infty} H_n \to \infty$, and the order of growth of H_n is given by $\lim_{n\to\infty} \frac{H_n}{\log n} = 1$. But much better estimates are known (see [2]), such as

$$H_n \;=\; \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} + O\left(\frac{1}{n^6}\right),$$

Here $O\left(\frac{1}{n^6}\right)$ denotes a function $f(n)$ such that $f(n) \le c\frac{1}{n^6}$ holds for some constant c.

where $\gamma \approx 0.5772$ is "Euler's constant."

Estimating factorials — Stirling's formula

The same method applied to

$$\log(n!) \;=\; \log 2 + \log 3 + \ldots + \log n \;=\; \sum_{k=2}^{n} \log k$$

yields

$$\log((n-1)!) \;<\; \int_1^n \log t\, dt \;<\; \log(n!),$$

where the integral is easily computed:

$$\int_1^n \log t\, dt \;=\; \Big[t\log t - t\Big]_1^n \;=\; n\log n - n + 1.$$

Thus we get a lower estimate on $n!$

$$n! \;>\; e^{n\log n - n + 1} \;=\; e\Big(\frac{n}{e}\Big)^n$$

and at the same time an upper estimate

$$n! \;=\; n\,(n-1)! \;<\; ne^{n\log n - n + 1} \;=\; en\Big(\frac{n}{e}\Big)^n.$$

Here a more careful analysis is needed to get the asymptotics of $n!$, as given by *Stirling's formula*

$$n! \;\sim\; \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

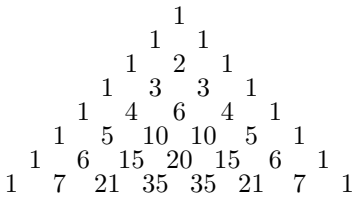
Here $f(n) \sim g(n)$ means that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

And again there are more precise versions available, such as

$$n! \;=\; \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{5140n^3} + O\left(\frac{1}{n^4}\right)\right).$$

Estimating binomial coefficients

Just from the definition of the binomial coefficients $\binom{n}{k}$ as the number of k-subsets of an n-set, we know that the sequence $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ of binomial coefficients



Pascal's triangle

- sums to $\sum_{k=0}^{n} \binom{n}{k} = 2^n$
- is symmetric: $\binom{n}{k} = \binom{n}{n-k}$.

From the functional equation $\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$ one easily finds that for every n the binomial coefficients $\binom{n}{k}$ form a sequence that is symmetric and *unimodal*: it increases towards the middle, so that the middle binomial coefficients are the largest ones in the sequence:

$$1 = \binom{n}{0} < \binom{n}{1} < \cdots < \binom{n}{\lfloor n/2 \rfloor} = \binom{n}{\lceil n/2 \rceil} > \cdots > \binom{n}{n-1} > \binom{n}{n} = 1.$$

Here $\lfloor x \rfloor$ resp. $\lceil x \rceil$ denotes the number x rounded down resp. rounded up to the nearest integer.

From the asymptotic formulas for the factorials mentioned above one can obtain very precise estimates for the sizes of binomial coefficients. However, we will only need very weak and simple estimates in this book, such as the following: $\binom{n}{k} \leq 2^n$ for all k, while for $n \geq 2$ we have

$$\binom{n}{\lfloor n/2 \rfloor} \;\geq\; \frac{2^n}{n},$$

with equality only for $n = 2$. In particular, for $n \geq 1$,

$$\binom{2n}{n} \;\geq\; \frac{4^n}{2n}.$$

This holds since $\binom{n}{\lfloor n/2 \rfloor}$, a middle binomial coefficient, is the largest entry in the sequence $\binom{n}{0}+\binom{n}{n}, \binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{n-1}$, whose sum is 2^n, and whose average is thus $\frac{2^n}{n}$.

On the other hand, we note the upper bound for binomial coefficients

$$\binom{n}{k} \;=\; \frac{n(n-1)\cdots(n-k+1)}{k!} \;\leq\; \frac{n^k}{k!} \;\leq\; \frac{n^k}{2^{k-1}},$$

which is a reasonably good estimate for the "small" binomial coefficients at the tails of the sequence, when n is large (compared to k).

References

[1] P. ERDŐS: *Beweis eines Satzes von Tschebyschef,* Acta Sci. Math. (Szeged) **5** (1930-32), 194-198.

[2] R. L. GRAHAM, D. E. KNUTH & O. PATASHNIK: *Concrete Mathematics. A Foundation for Computer Science,* Addison-Wesley, Reading MA 1989.

[3] G. H. HARDY & E. M. WRIGHT: *An Introduction to the Theory of Numbers,* fifth edition, Oxford University Press 1979.

[4] P. RIBENBOIM: *The New Book of Prime Number Records,* Springer-Verlag, New York 1989.

http://www.springer.com/978-3-642-00855-9

Proofs from THE BOOK
Aigner, M.; Ziegler, G.M.
2010, VIII, 274 p. 250 illus., Hardcover
ISBN: 978-3-642-00855-9