Apprentice Program

Instructor: Lászlo Babai
Notes taken by Matt Holden and Kate Ponto
NOT PROOF-READ

June 30,2006

Chebyshev's Theorem

Recall that $\pi(x)$ is the number of primes less than or equal to x. The goal of this lecture is to prove:

Theorem 1 (Chebyshev's Theorem).

$$\pi(x) = \Theta\left(\frac{x}{\ln(x)}\right) \tag{1}$$

there exist $c_1, c_2 > 0$ *such that for all* $x \geq 2$*, we have* $c_1 \frac{x}{\ln(x)} < \pi(x) < c_2 \frac{x}{\ln(x)}$.

This is a weakened version of the prime number theorem and it follows from results about binomial coefficients.

1. $\binom{2n}{n} < 4^n = 2^{2n}$

   The number of subsets of size n in a set with $2n$ elements is $\binom{2n}{n}$. The number of subsets of a set with $2n$ elements is 2^{2n}. Since the subsets of size n are contained in all of the subsets of the set with $2n$ elements we see that $\binom{2n}{n} < 4^n = 2^{2n}$.

   This same reasoning shows that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{2}$$

since both sides are the number of subsets of a set with n elements. Therefore,

$$\binom{2n}{n} < \sum_{k=0}^{2n} \binom{2n}{k} = 2^{2n} \tag{3}$$

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

   The maximum of $\binom{2n}{k}$ as a function of k is $\binom{2n}{n}$, so $\binom{2n}{n}$ is the largest element in $\sum_{k=0}^{2n} \binom{2n}{k} = 2^{2n}$. Taking the average of the elements in this sum we have

   $$\frac{2^{2n}}{2n+1} = \frac{\sum \binom{2n}{k}}{2n+1} < \binom{2n}{n} \tag{4}$$

   since the average is always smaller than the largest value.

3. $\binom{2n+1}{n} < 4^n = 2^{2n}$

   The maximum of $\binom{2n+1}{k}$ as a function of k is $\binom{2n+1}{n} = \binom{2n+1}{n+1}$ (recall $\binom{n}{k} = \binom{n}{n-k}$). Then

   $$2\binom{2n+1}{n} = \binom{2n+1}{n} + \binom{2n+1}{n+1} < \sum_{k=1}^{2n+1} \binom{2n+1}{k} = 2^{2n+1} \tag{5}$$

   and so $\binom{2n+1}{n} < 4^n$.

What is the largest power of 7 that divided (1000!)? How many multiples of 7 appear in $1, 2, 3, \ldots, 1000$?

More generally, find the largest ℓ such that $p^\ell \,\big|\, n!$. The number of multiples of p among $1, \ldots n$ is $\lfloor \frac{n}{p} \rfloor$. So

$$\ell = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \ldots \tag{6}$$

since we added 1 for each multiple of p, and another 1 for each multiple of p^2, ... Writing this in summation notation,

$$\ell = \sum_{s=1}^{\infty} \left\lfloor \frac{n}{p^s} \right\rfloor = \sum_{s=1}^{\lfloor \log_p n \rfloor} \left\lfloor \frac{n}{p^s} \right\rfloor . \tag{7}$$

We can replace ∞ by $\lfloor \log_p n \rfloor$ since $s > \lfloor \log_p n \rfloor$ implies $\left\lfloor \frac{n}{p^s} \right\rfloor = 0$. We can also find an upper bound for ℓ:

$$\ell < \frac{n}{p} + \frac{n}{p^2} + \frac{n}{p^3} + \ldots = \frac{n}{p} \cdot \frac{1}{1 - \frac{1}{p}} = \frac{n}{p-1}. \tag{8}$$

Digression: $\lfloor \frac{1000000}{7} \rfloor = 142857$.

$$\begin{array}{rcl} 142857 \times 2 & = & 285714 \\ 142857 \times 3 & = & 428571 \\ 142857 \times 4 & = & 571428 \\ 142857 \times 7 & = & 999999 \end{array}$$

The last equality means that $\frac{1}{7} = 0.\dot{1}4285\dot{7}$.

Multiplying 142857 by $2, 3, 4, 5$, or 6 gives a cyclic permutation of the digits. Something similar happens for 17, but not for $3, 5, 11$, or 13.

Exercise 2. If p is a prime $\neq 2$ or 5, then the length of the period of $\frac{1}{p}$ divides $p - 1$.

Exercise 3. If the period of $\frac{1}{p}$ is $p-1$ then multiplying the first $p-1$ elements of the decimal expansion of $\frac{1}{p}$ by $1, 2, \ldots p-1$ gives all cyclic permutations of that number.

Exercise 4. If A is a 6 digit number and $A, 2A, 3A, \ldots, 6A$ have the same digits as A, then $A = 142857$.

End of Digression

Theorem 5. *If $p^t \bigm| \binom{n}{k}$ then $p^t \leq n$.*

If $p \bigm| \binom{n}{k}$ then $p \leq n$ since $\binom{n}{k} = \frac{n!}{k!(n-k!)}$ and no prime larger than n divides $n!$.

Proof of the Theorem. If t is the largest integer such that $p^t \bigm| \binom{n}{k}$ then from formula (7) we get

$$t = \sum_{s=1}^{\infty} \left(\left\lfloor \frac{n}{p^s} \right\rfloor - \left\lfloor \frac{k}{p^s} \right\rfloor - \left\lfloor \frac{n-k}{p^s} \right\rfloor \right). \tag{9}$$

Some sample calculations of $\lfloor \frac{n}{p^s} \rfloor - \lfloor \frac{k}{p^s} \rfloor - \lfloor \frac{n-k}{p^s} \rfloor$ with $n = 1000$, $k = 73$, $p = 5$:

s	$\lfloor \frac{n}{p^s} \rfloor - \lfloor \frac{k}{p^s} \rfloor - \lfloor \frac{n-k}{p^s} \rfloor$
1	$200 - 14 - 185 = 1$
2	$40 - 2 - 37 = 1$
3	$8 - 0 - 7 = 1$
4	$1 - 0 - 1 = 0$

Exercise 6. Show that

$$\left\lfloor \frac{\lfloor \frac{n}{p} \rfloor}{p} \right\rfloor = \left\lfloor \frac{n}{p^2} \right\rfloor \tag{10}$$

Exercise 7. For all $x, y \in \mathbb{R}$, $0 \leq \lfloor x + y \rfloor - \lfloor x \rfloor - \lfloor y \rfloor \leq 1$.

Exercise 7 implies that $0 \leq \lfloor \frac{n}{p^s} \rfloor - \lfloor \frac{k}{p^s} \rfloor - \lfloor \frac{n-k}{p^s} \rfloor \leq 1$. So t is less than or equal to the number of terms in the sum, $t \leq \lfloor \log_p n \rfloor$, and

$$p^t \leq p^{\lfloor \log_p n \rfloor} \leq p^{\log_p n} = n. \tag{11}$$

$\square$

Theorem 8. *For all positive real numbers x, $\prod_{p \leq x} p \leq 4^x$.*

Observation: It suffices to prove this for positive integers x, then it holds for all positive reals.

Proof. This proof is by induction on x.

Base case: $x = 0$, $\prod_{p \leq 0} p = 1 \leq 4^0 = 1$ (empty products); $x = 1$, $\prod_{p \leq 1} p = 1 \leq 4^1 = 4$; $x = 2$, $\prod_{p \leq 2} p = 2 \leq 4^2 = 16$

Induction step: Assume $x \geq 3$. Assume $\prod_{p \leq y} p \leq 4^y$ for all $y < x$. If x is even $\prod_{p \leq x} p = \prod_{p \leq x-1} \leq 4^{x-1} < 4^x$.

If x is odd, say $x = 2y + 1$,

$$\prod_{p \leq 2y+1} p = \left(\prod_{p \leq y+1} p \right) \left(\prod_{y+2 \leq p \leq 2y+1} p \right). \tag{12}$$

Let $A = \prod_{y+2 \leq p \leq 2y+1} p$.

Lemma 9. $A \Big| \binom{2y+1}{y}$

Then $A \Big| \binom{2y+1}{y} < 4^y$. By the induction hypothesis $\prod_{p \le y+1} p \le 4^{y+1}$ so

$$\prod_{p \le 2y+1} p \le 4^{y+1} 4^y = 4^{2y+1} = 4^y. \tag{13}$$

□

Proof of the Lemma. $\binom{2y+1}{y} = \frac{(2y+1)!}{y!(y+1)!}$ The primes in A divide the numerator and not the denominator of $\frac{(2y+1)!}{y!(y+1)!}$ and so they divide $\binom{2y+1}{y}$. □

This completes the proof of Theorem 8: $\prod_{p \le x} p \le 4^x$.

Since $p \ge 2$ for every prime p, it follows that

$$2^{\pi(x)} \le \prod_{p \le x} p \le 4^x, \tag{14}$$

which gives the bound $\pi(x) \le 2x$. This bound is not very good, because replacing p with 2 is not a very good estimate. Instead, let's try using $\sqrt{x}$:

$$4^x \ge \prod_{p \le x} p \ge \prod_{\sqrt{x} \le p \le x} p \ge \sqrt{x}^{\pi(x) - \pi(\sqrt{x})}. \tag{15}$$

Taking base-2 logarithms yields

$$2x \ge \left(\pi(x) - \pi(\sqrt{x})\right) \cdot \frac{1}{2} \log_2 x, \tag{16}$$

which implies

$$\pi(x) \le \pi(\sqrt{x}) + \frac{4x}{\log_2 x} \le \sqrt{x} + \frac{4x}{\log_2 x}. \tag{17}$$

But $\sqrt{x}$ is small compared to $x/\log x$:

Exercise 10. Use calculus to show that $\sqrt{x} = o(x/\log x)$.

Thus, equation (17) yields

$$\pi(x) \le \sqrt{x} + \frac{4x}{\log_2 x} \sim \frac{4x}{\log_2 x}, \tag{18}$$

so we obtain the asymptotic inequality

$$\pi(x) \lesssim c \frac{x}{\ln x}, \tag{19}$$

for the value $c = 4 \ln 2$. This implies

Theorem 11 (Chebyshev's upper bound).

$$\pi(x) < c' \frac{x}{\ln x} \tag{20}$$

for any $c' > 4\ln 2$ *and all sufficiently large* x.

We next want to prove:

Theorem 12 (Chebyshev's lower bound). *There exists* $c > 0$ *such that* $\pi(x) \gtrsim c\frac{x}{\ln x}$.

To prove this, consider the prime factorization

$$\binom{2n}{n} = \prod_{p \le 2n} p^{k_p}, \tag{21}$$

for some integers $k_p \ge 0$. Then

$$\frac{4^n}{2n+1} < \binom{2n}{n} = \prod_{p \le 2n} p^{k_p} \le (2n)^{\pi(2n)}, \tag{22}$$

where the first inequality follows from equation (4) and the second follows from the fact that $p^{k_p} \le 2n$. Taking logarithms yields

$$\pi(2n) \cdot \log_2(2n) > n \log_2 4 - \log_2(2n+1) \sim n \log_2 4 = 2n, \tag{23}$$

which implies

$$\pi(2n) \gtrsim \frac{2n}{\log_2(2n)}. \tag{24}$$

This proves Theorem 12 for even n.

Exercise 13. Finish the proof by extending the result to all n.

This proof of Chebyshev's estimate is due to Paul Erdös, who used the same ideas to give an elementary proof of Bertrand's Postulate: for every n, there is a prime p such that $n \le p < 2n$. It is an open question whether, for every n, there is a prime p such that $n^2 < p < (n+1)^2$. An affirmative answer to this question would imply the famous Riemann Hypothesis, which we now briefly discuss.

Consider the *zeta function* defined by the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{25}$$

which converges for $s > 1$. In fact, this converges for all complex numbers $s \in \mathbb{C}$ such that $\operatorname{Re} s > 1$. Using techniques from complex analysis, we can then extend the function ζ to the entire complex plane $\mathbb{C}$, with the exception of $s = 1$. Riemann's hypothesis says that if $0 < \operatorname{Re} s < 1$ and $\zeta(s) = 0$ then $\operatorname{Re} s = \frac{1}{2}$. A proof of the Riemann hypothesis would give us better estimates of $\pi(x)$.

The Prime Number Theorem says $\pi(x) \sim \frac{x}{\ln x}$. In fact, a better approximation (known to Gauss) is

$$\pi(x) \sim \operatorname{li}(x) = \int_2^x \frac{dt}{\ln t}, \tag{26}$$

and we would like to estimate the error term $|\pi(x) - \operatorname{li}(x)|$. For example, we have

$$\left| \frac{x}{\ln x} - \operatorname{li}(x) \right| = \Theta\left(\frac{x}{(\ln x)^2} \right). \tag{27}$$

The Riemann hypothesis is equivalent to the error estimate

$$|\pi(x) - \operatorname{li}(x)| = O(\sqrt{x}). \tag{28}$$

It is known that there exists an $\epsilon > 0$ such that $|\pi(x) - \operatorname{li}(x)| < x^{1-\epsilon}$.

Quadratic residues

Our goal is to prove the following:

Theorem 14. *If p is prime and $p \equiv 1 \pmod 4$ then $p = a^2 + b^2$ for some integers a, b.*

Recall that a is a *quadratic residue* (mod p) if $x^2 \equiv a \pmod p$ for some x. If no such x exists then a is a quadratic non-residue. By convention, if $a \equiv 0 \pmod p$ then a is neither a quadratic residue nor a non-residue.

Proposition 15. *If p is an odd prime then the number of quadratic residues in $\{1, 2, \ldots, p-1\}$ is $\frac{p-1}{2}$.*

As a corollary, we see that there are also $\frac{p-1}{2}$ quadratic non-residues.

Proof. Notice that $(p-x)^2 \equiv (-x)^2 \equiv x^2 \pmod{p}$, so the number of quadratic residues is *at most* $\frac{p-1}{2}$. Next, suppose $x^2 \equiv y^2 \pmod{p}$. Then $p \mid (x^2-y^2) = (x+y)(x-y)$, which implies that $p \mid (x+y)$ or $p \mid (x-y)$, hence $x \equiv -y \pmod{p}$ or $x \equiv y \pmod{p}$. Thus, we have shown that $x^2 \equiv y^2 \pmod{p}$ iff $x \equiv \pm y \pmod{p}$, which proves our claim. $\square$

Observe that we can find a quadratic residue of p simply by squaring an integer mod p, but finding a quadratic non-residue is more difficult. However, the proposition tells us that an integer chosen at random from $\{1, 2, \ldots, p\}$ will be a quadratic non-residue with probability 1/2. Thus, the probability that k random choices produces no quadratic non-residue is 2^{-k}, and the expected number of choices needed to find a non-residue is 2.

Question: when is -1 a quadratic residue mod p? In other words, for which p does there exist x such that $x^2 \equiv -1 \pmod{p}$. Well, $p \mid (x^2+1)$ implies $p \equiv 1 \pmod{4}$. We will see that the converse also holds: if $p \equiv 1 \pmod{4}$ then -1 is a quadratic residue mod p.

Notation: the *Legendre symbol* is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & , \quad a \text{ is a quadratic residue,} \\ -1 & , \quad a \text{ is a quadratic nonresidue,} \\ 0 & , \quad p \mid a. \end{cases} \tag{29}$$

Exercise 16. Prove that the Legendre symbol is multiplicative: for every prime p and integers a, b,

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{30}$$

Exercise 17. If $p \nmid a$ then $a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}$.

Exercise 18. If $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$ then $\left(\frac{a}{p}\right) = -1$.

Theorem 19. *If* $\left(\frac{a}{p}\right) = -1$ *then* $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$.

The above exercises and the previous theorem imply the following result:

Theorem 20 (Euler). $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$.

Definition 21. g is a *primitive root* mod p if $\{1, g, g^2, \ldots, g^{p-2}\}$ are all nonzero residues mod p. In other words, for all b, if $p \nmid b$ then $b \equiv g^j \pmod{p}$ for some j.

Exercise 22. Check that 10 is a primitive root mod 7, but 2 is not.

Theorem 23. *For all primes p, there is a primitive root mod p.*

Exercise 24. Use Theorem 23 to prove Euler's formula for the Legendre symbol (Theorem 20).