UNSW AUSTRALIA.
SCHOOL OF MATHEMATICS AND STATISTICS.

MATH5645: TOPICS IN NUMBER THEORY.
§2 SOME ELEMENTARY RESULTS ON THE DISTRIBUTION OF PRIMES:

The oldest result concerning the prime numbers is probably the fact that they increase without bound. This goes back to Euclid.

Theorem 2.1: There are infinitely many primes.

Proof 1: (Euclid) Suppose not, then we can write down the set S of **all** primes,

$$S = \{p_1, p_2, p_3, \cdots, p_n\}.$$

Now let $N = p_1 p_2 \cdots p_n + 1$. If N is prime we have a contradiction since clearly $N \notin S$. Hence N has a prime factor q, but $q \notin S$ and again we have a contradiction.

Proof 2: (Kummer). Again, suppose not and again let S be the set of all primes as above.

Proof 3:

(Proof 3 is constructive and not by contradiction. It allows you to explicitly write down a number with at least 1000 distinct prime factors!).

Definition: Let p_n denote the nth prime (so $p_1 = 2, p_2 = 3, p_4 = 5$ and so on) and let $\pi(n)$ denote the number of primes less or equal to n.

Thus $p_k = n \Leftrightarrow \pi(n) = k$.

Ex: $p_{10} = 29$ and $\pi(29) = 10$.

From Euclid's proof we can easily deduce that $p_n \leq p_1 p_2 \cdots p_{n-1} + 1$. Using this we can prove:

Theorem 2.2: $p_n \leq 2^{2^{(n-1)}}$, for $n \geq 1$.

Proof: (By induction).

Corollary: $\pi(x) \geq \frac{\log\log(x)}{\log 2}$.

Proof:

Note that this also shows there are infinitely many primes, but as a bound it is rather poor. For example, it says that $\pi(10^9) \geq \log\log 10^9 / \log 2 \approx 4$ (!)

Primes and Squares:

How do the primes compare with the squares? There are infinitely many of both, but since $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, in some sense there are not too many squares. Does $\sum_{n=1}^{\infty} \frac{1}{p_n}$ converge? We shall presently prove that this series diverges, so that in some sense, there are more primes than squares. It is interesting to note that if we take the above sum over all **known** primes, then the sum is relatively small!

Before we tackle the proof of this result, I want to introduce one of the key ideas (due to Euler) in the analysis of prime numbers.

Euler's Product:

Ignoring problems of convergence, let us see what we get when we expand

$$\left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots\right)\left(1 + \frac{1}{3} + \frac{1}{3^2} + \cdots\right)\left(1 + \frac{1}{5} + \frac{1}{5^2} + \cdots\right)\cdots$$

$$= \prod_{p \text{ prime}} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right).$$

One the one hand, we have $\left(1+\frac{1}{p}+\frac{1}{p^2}+\cdots\right)=\left(1-\frac{1}{p}\right)^{-1}$ and so the product can be written as

$$\prod_{p \text{ prime}} \left(1-\frac{1}{p}\right)^{-1}.$$

On the other hand, Euler noted that by unique factorisation in the integers, the expansion of the left hand side above, produces the reciprocal of each positive integer **exactly** once. So

$$\prod_{p \text{ prime}} \left(1-\frac{1}{p}\right)^{-1} = \sum_{k=1}^{\infty} \frac{1}{k}.$$

Now at this stage, you will object that this is nonsense, since the right-hand side (and in fact the left hand side) diverges. Nonetheless, this 'equation' is really trying to say something important. We can use the same argument to prove that

$$\prod_{p \text{ prime}} \left(1-\frac{1}{p^s}\right)^{-1} = \sum_{k=1}^{\infty} \frac{1}{k^s},$$

which **is** valid for $s>1$. The function $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^s}$, denoted by $\zeta(s)$, plays a central role in analytic number theory and is known as the *Riemann Zeta Function.* We will study this function in much more detail later. For the moment, let us simply observe that

$$\prod_{i=1}^{\pi(n)} \left(1-\frac{1}{p_i}\right)^{-1} = 1+\frac{1}{2}+\cdots+\frac{1}{n}+\ldots > 1+\frac{1}{2}+\cdots+\frac{1}{n}.$$

The Sum $\displaystyle\sum_{p\leq x} \frac{1}{p}$.

We can now prove that the sum $\displaystyle\sum_{p\leq x} \frac{1}{p}$ diverges as $x\to\infty$ and furthermore, obtain a good estimation as to how fast it grows.

Theorem 2.3a: $\sum_{i=1}^{\infty} \frac{1}{p_i}$ diverges.

Proof: (There are many proofs of this. The proof given here, although slightly longer than some others contains a number of ideas which we will use again later.)

Although this series diverges, it does so very *slowly*. We would like to get some idea of the rate of growth. The following result gives a lower bound on $\sum_{p \leq x} \frac{1}{p}$, but from it, one can show that the true order of magnitude of this sum is $\log \log x$. As usual, in the statement $p \leq x$, it is assumed that p is prime.

As often, we need a simple Lemma:

Lemma:

(i) For $x \geq 1$, $\quad \sum_{n \leq x} \frac{1}{n} \geq \log x$.

(ii) For $x \geq 1$, $\quad 1 + \frac{1}{x} \leq e^{\frac{1}{x}}$.

Proof: (i) is left as an exercise - draw a diagram!

(ii) Follows immediately from the series expansion of the right-hand side.

Theorem 2.3b

For $x > 2$,

$$\sum_{p \leq x} \frac{1}{p} \geq \log \log x - \log \zeta(2).$$

Proof: (Again there are a number of proofs of this. I have chosen the one I think is the nicest.)

This, of course, gives yet another proof that the series $\sum_p \frac{1}{p}$ diverges. We will look again in more detail at this sum in Chapter 6.

Bertrand's Postulate:

We now state and prove an important result, known as *Bertand's Postulate*. (It is now a **Theorem**, but the name comes from the time when it was a conjecture.) It was first proven by Chebychev (1852) but the proof given here is a modification of a proof due to Erdös.

Lemma:
(a) For all real $x \geq 2$, $\prod_{p \leq x} p \leq 4^{x-1}$.
(b) (Legendre's Theorem) The number $n!$ contains the prime factor p exactly $\sum_{k \geq 1} \lfloor \frac{n}{p^k} \rfloor$ times.
(c) For $n \geq 1$, $\binom{2n}{n} \geq \frac{4^n}{2n+1}$

Proof:
(a) See tutorial problems.

(b) Well known. (For example, find the number of zeros at the end of 1000!)

(c) $\binom{2n}{n}$ is the greatest among the co-efficients $\binom{2n}{0}, \binom{2n}{1}, \cdots, \binom{2n}{2n}$. Hence $\binom{2n}{n}$ cannot be less than their average which is $\frac{1}{2n+1} 2^{2n} = \frac{4^n}{2n+1}$.

Theorem 2.4: (Bertrand's Postulate).

If $n > 2$ then there exists a prime p satisfying $n < p \leq 2n$.

Proof:
If $n < 600$ then the result is true, since we only need to check that
$2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 547$ are all primes.

We proceed by trying to find a bound on the size of $\binom{2n}{n}$, using its prime factors. These prime factors

fall into 4 regions. For example if $n = 39$,

$$\binom{78}{39} = \begin{array}{c|c|c|c} 2^4.5^2.7^2 & 11.23 & & 41.43.47.53.59.61.67.71.73 \\ \hline p \le \sqrt{2n} & \sqrt{2n} < p \le \frac{2}{3}n & \frac{2}{3}n < p \le n & n < p < 2n \end{array}$$

Let $r(p)$ denote the largest power of a prime p which divides $\binom{2n}{n}$. In the example above, with $n = 39$, we have $r(2) = 4, r(71) = 1$ and $r(29) = 0$

Using Legendre's theorem, $r(p) = \sum_{k\ge 1} \lfloor \frac{2n}{p^k} \rfloor - 2\lfloor \frac{n}{p^k} \rfloor$.

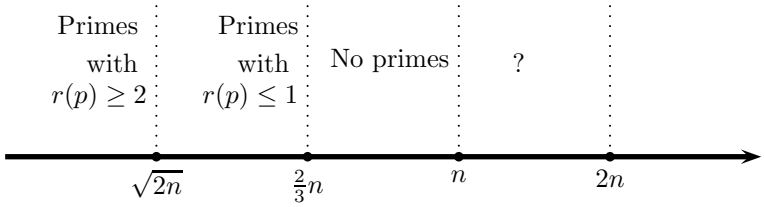
Each summand satisfies,

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

and so is at most 1. Thus, $r(p) \le \max\{t : p^t \le 2n\}$, (since if $p^t > 2n$, the summand is vacuous). So in particular, in the region of all primes $p \le \sqrt{2n}$, the contribution to $\binom{2n}{n}$ is $\prod_{p\le\sqrt{2n}} p^{r(p)} \le \prod_{p\le\sqrt{2n}} 2n$.

Also if $r(p) \ge 2$, we have $p \le \sqrt{2n}$. Thus, primes p greater than $\sqrt{2n}$ appear at most once in the factorisation of $\binom{2n}{n}$.

Now for a rather striking observation (which is the key to the whole thing). Primes p, for which $\frac{2}{3}n < p \le n$, do NOT divide $\binom{2n}{n}$ at all! For if $3p > 2n$, p and $2p$ are the only multiples of p in the numerator of $\frac{(2n)!}{n!n!}$ while $p|n!$ (in fact $p||n!$) and so the two p's on the top cancel with the two p's on the bottom.



We can now estimate the size of $\binom{2n}{n}$. For $n \ge 3$, using the Lemma,

$$\frac{4^n}{2n+1} \le \binom{2n}{n} \le \prod_{p\le\sqrt{2n}} 2n \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p \prod_{n<p\le 2n} p$$

Now suppose that the desired result is **false**, then the last product is 1, so we have

$$\frac{4^n}{2n+1} \le (2n)^{\sqrt{2n}} \prod_{\sqrt{2n}<p\le\frac{2}{3}n} p \le (2n)^{\sqrt{2n}} \prod_{p\le\frac{2}{3}n} p \le (2n)^{\sqrt{2n}} 4^{\frac{2}{3}n-1} < (2n)^{\sqrt{2n}} 4^{\frac{2}{3}n}$$

using part (a) of the Lemma.
This last inequality implies,

$$4^{\frac{1}{3}n} \le (2n+1)(2n)^{\sqrt{2n}} < (2n)^2(2n)^{\sqrt{2n}} = (2n)^{2+\sqrt{2n}}.$$

Taking logs this is equivalent to $\frac{1}{3}n\log 4 \le (\sqrt{2n}+2)\log(2n)$ which is false for sufficiently large n. In fact a MAPLE sketch indicates that it is false for $n > 600$.

Notes:

1. The above result can also be written as $\pi(2n) - \pi(n) \ge 1$ for $n \ge 2$. This can be generalised to show that $\pi(2n) - \pi(n) \ge 2$ for $n \ge 6$.

2. Using the same ideas above, it can be shown that $\prod_{n<p\leq 2n} p \geq 2^{\frac{n}{30}}$ for $n \geq 4000$, so there are at least $\log_{2n} 2^{\frac{n}{30}}$ primes between n and $2n$. The Prime Number Theorem gives the true order to be $\frac{n}{\log n}$.

3. Is there always a prime between n^2 and $(n+1)^2$? This is unsolved.

Primes in Arithmetic Sequences:

We saw earlier that the primes naturally split into those which are congruent to 1 mod 4 and those congruent to -1 mod 4. (For example the 1 mod 4 primes are the sum of two squares.)

Theorem 2.5a:

There are infinitely many primes congruent to -1 mod 4.

Proof: We model the proof on that of Theorem 2.1.

Note that the same kind of argument will NOT work for primes congruent to 1 mod 4.

Theorem 2.5b: There are infinitely many primes congruent to 1 mod 4.

Proof:

In the tutorial problems, you will be asked to prove similar special cases. All of these are special cases of the following key result due to Dirichlet, namely, if $(a, b) = 1$ then $\{a + bn : n = 1, 2, \cdots\}$ contains infinitely many primes. The *ad hoc* methods used above will not give the general result. We need a more sophisticated approach to the problem, and the proof of Dirichlet's Theorem will be covered in Chapter 4.

Chebychev's Bounds:

Let us return to the function $\pi(x)$. At the age of 14 (!!), Gauss guessed that this function can be asymptotically approximated by $\frac{x}{\log x}$ in the sense that $\frac{\pi(x)}{\frac{x}{\log x}} \to 1$ as $x \to \infty$. This is known as the **prime number theorem**. It will be the other main theorem we will prove in this course. Gauss could not prove it, but Chebychev was able to show that it was at least roughly correct. More precisely, he proved that **if** $\pi(x) \sim \frac{Cx}{\log x}$ then $C = 1$ and that there exist constants A and B such that

$$\frac{Ax}{\log x} \leq \pi(x) \leq \frac{Bx}{\log x}.$$

We will find such constants, but not quite as good as those which Chebychev found.

To progress to the result, we need a couple of technical facts about the binomial co-efficients.

Lemma:

(i) For $n \geq 1$, $2^n \leq \left(\begin{array}{c} 2n \\ n \end{array} \right) < 2^{2n}$

(ii) For $n \geq 1$, $p \left| \left(\begin{array}{c} 2n \\ n \end{array} \right) \right.$ for **every** prime p such that $n < p \leq 2n$.

Proof:

We can now see that Gauss' guess was in the right ballpark.

Theorem 2.6 (Chebychev)

$$\frac{2}{3}\frac{x}{\log x} < \pi(x) < 1.7\frac{x}{\log x}.$$

(Chebychev in fact was able to reduce the left hand constant to 0.921 and the right hand one to 1.105, but these can be further refined.)

Proof:

Firstly we note that $\pi(x) < 1.7\frac{x}{\log x}$ for $x < 1200$ by trial and error (or use MAPLE).

Now suppose the right-hand inequality is true for all $x \leq n$.
Consider $\left(\begin{array}{c} 2n \\ n \end{array} \right)$. From the Lemma, we have

$$\prod_{n<p\leq 2n} p \leq \left(\begin{array}{c} 2n \\ n \end{array} \right) < 2^{2n}.$$

Now there are $\pi(2n) - \pi(n)$ primes (strictly) between n and $2n$ each greater than n so

$$n^{\pi(2n)-\pi(n)} \leq \prod_{n<p\leq 2n} p < 2^{2n}.$$

Taking logarithms and dividing by $\log n$ we have

$$\pi(2n) - \pi(n) < \frac{2n\log 2}{\log n} < 1.39\frac{n}{\log n}$$

(NB. $2\log 2 \approx 1.386$).
Now by assumption $\pi(n) < 1.7\frac{n}{\log n}$ so

$$\pi(2n) < 1.39\frac{n}{\log n} + \pi(n) < (1.39 + 1.7)\frac{n}{\log n} = 3.09\frac{n}{\log n}.$$

It is left as an exercise to show that $\frac{3.09n}{\log n} < 1.7 \times \frac{2n}{\log(2n)}$ if $n > 1200$. (In fact if $n > 1001$.)

Hence, if the inequality is true for n then is it true for $2n$. Also

$$\pi(2n+1) \leq \pi(2n) + 1 < 3.09\frac{n}{\log n} + 1 < 1.7 \times \frac{(2n+1)}{\log(2n+1)}$$

if $n > 1200$ (This last inequality is also left as an exercise but can be easily seen if we use MAPLE to draw the relevant graphs.)

Hence, if the inequality is true for n then it is true for $2n$ and $2n+1$, and so holds for all n.

Now for the left-hand side.

By Legendre's Theorem, (in the earlier Lemma), the highest power of p which divides $\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}$ is given by p^{ν_p} where

$$\nu_p = \sum_r \left(\lfloor \frac{n}{p^r} \rfloor - \lfloor \frac{k}{p^r} \rfloor - \lfloor \frac{n-k}{p^r} \rfloor \right).$$

Now it is an exercise to show that $\lfloor x \rfloor - \lfloor y \rfloor \leq \lfloor x - y \rfloor + 1$ and so each term in the sum is either 0 or 1. Furthermore, if $p^r > n$, i.e. if $r > \log_p n$, then each term of the sum is 0. Thus, $\nu_p \leq \log_p n$, and so $p^{\nu_p} \leq p^{\log_p n} = n$. Hence we have

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \prod_{p|\left(\begin{smallmatrix} n \\ k \end{smallmatrix} \right)} p^{\nu_p} \leq n^{\pi(n)},$$

since there are at most $\pi(n)$ terms in the product.
Also we note that $2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k} \leq (n+1)n^{\pi(n)}$ (there are $(n+1)$ terms in the sum) and so $n^{\pi(n)} \geq \frac{2^n}{n+1}$.
This gives

$$\pi(n) \geq \frac{1}{\log n}(n \log 2 - \log(n+1)) > \frac{2}{3}\frac{n}{\log n}$$

if $n > 220$. This last inequality is an exercise. (Note that $\log 2 \approx \frac{2}{3}$). (A MAPLE plot here is convincing).

Theorem 2.7: For $n \geq 1$, the nth prime, p_n, satisfies the inequalities

$$0.58n \log n < p_n < 3n(\log n + \frac{1}{2}).$$

Proof: If $k = p_n$ then $k \geq 2$ and $n = \pi(k)$.
Applying the previous theorem, we have

$$n = \pi(k) < \frac{1.7k}{\log k} = \frac{1.7p_n}{\log p_n}.$$

Hence $p_n > 0.58n \log p_n > 0.58n \log n$.

Also, $n = \pi(k) > \frac{2}{3}\frac{k}{\log k} = \frac{2}{3}\frac{p_n}{\log p_n}$, and hence

$$p_n < \frac{3}{2}n \log p_n \qquad (*).$$

Now $\log x < \sqrt{x}$, if $x > 1$, so $\log p_n \leq \sqrt{p_n}$. Thus, using (*), $p_n < \frac{3}{2}n\sqrt{p_n}$ giving $p_n < \frac{9}{4}n^2$. Taking logs, and using (*) again, we obtain

$$p_n < 3n(\log \frac{3}{2}n) < 3n(\log n + \frac{1}{2}), \quad (\text{ since } \log \frac{3}{2} \approx 0.4).$$

Ex:

n	$0.58n \log n$	p_n	$3n(\log n + \frac{1}{2})$
3	1.91	5	14.4
4	3.22	7	22.6
9	11.47	23	72.8
10	13.35	29	84.1

Notes:
1. The 'correct' order of magnitude of p_n is $n \log n$. (This is equivalent to the Prime Number Theorem.)

2. We obtain yet another proof that $\sum \frac{1}{p}$ diverges, since we can compare this series to $\sum \frac{1}{n \log n}$.

3. These difficult, but *ad hoc*, methods, do not produce particularly satisfying results. We need much better techniques.