The Weak Prime Number Theorem

1 Introduction

Notation 1.1 $\pi(n)$ is the number of primes that are $\le n$.

The Prime Number Theorem states that $\pi(n)$ tends to $\frac{n}{\ln n}$ as n goes to infinity. (Formally the ratio of the two tends to 1.) Note that there are no hidden multiplicative constants– the theorem is tight. It was proven independently by Hadamard (1896) and de la Vallee Poussin (1896). An easy corollary of the Prime Number Theorem is Bertrand's Postulate: for all large n there is a prime between n and $2n$. (Its actually known that for $n \ge 3$ this is true.) This was proven by Chebyshev (1850). Bertrand's Postulate is used in Theoretical Computer Science since you often need to find a prime. We give two examples.

1. The proof that EQUALITY has Randomized Communication Complexity $O(\log n)$ uses that there exists a prime between n and $2n$.
2. FILL IN LATER

The Prime Number Theorem is difficult to prove. In this note we prove a weaker version of the Prime Number Theorem, due to Chebyshev (1850?), namely $\pi(n) = \Theta(\frac{n}{\ln n})$. We will do this by getting upper and lower bounds on $\pi(n)$. In both cases are constants are quite good. This version sufficient to obtain a weak version of Bertrand's Postulate. This weak version suffices for all computer science applications. Chebyshev also proved that if the ratio of $\pi(n)$ to $\frac{n}{\ln n}$ existed then it was 1.

Our approach here is to get really good constants but have the result hold for large n. Alternatively one can, using similar techniques, obtain a result that has less good constants, but holds for all n.

I do not know if the proof presented here is Chebyshev's proof. I doubt it since he got better constants (see Note 7.3).

2 Some Really Easy Theorem

Before presenting the Weak Prime Number Theorem we present some very elementary upper and lower bounds on $\pi(n)$.

Theorem 2.1

1. $(\forall n)[\pi(n) \ge \frac{1}{3}\log_2 n]$.

2. $(\forall n)[\pi(n) \le n + 1 - \log_2 n]$.

Proof:

Let $COMP$ be the number of composite numbers that are $\le n$.

a) Let $x \in COMP$ and $x \le n$. We factor x so that $x = p_1^{a_1} \cdots p_{a_{\pi(n)}}^{a_n}$. Write each $a_i = 2b_i + c_i$ where $c_i \in \{0,1\}$. Hence $x = p_1^{2b_1} p_2^{2p_2} p_3^{2b_3} \cdots p_{\pi(n))}^{2b_{\pi(n)}} p_1^{c_1} p_2^{c_2} \cdots p_{\pi(n)}^{c_{\pi(n)}}$. Note that this can be written as $m^2 p_1^{c_1} p_2^{c_2} \cdots p_{\pi(n)}^{c_{\pi(n)}}$. How many numbers are of this form?

There are at most $\sqrt{n}$ numbers of the form m^2 where $m^2 \le n$. There are at most $2^{\pi(n)}$ numbers of the form $p_1^{c_1} p_2^{c_2} \cdots p_{\pi(n)}^{c_{\pi(n)}}$ where each $c_i \in \{0,1\}$. Hence there

$$
\begin{aligned}
COMP \le & \quad \sqrt{n} 2^{\pi(n)} \\
n - \pi(n) \le & \quad \sqrt{n} 2^{\pi(n)} \\
n \le & \quad \sqrt{n} 2^{\pi(n)} + \pi(n)
\end{aligned}
$$

if $\pi(n) \le (\frac{1}{3}) \log_2 n$ then

$$n \le \sqrt{n} 2^{\pi(n)} + \pi(n) \le \sqrt{n} n^{1/3} + \frac{1}{3} \log_2 n \le n^{5/6} + \frac{1}{3} \log_2 n$$

which is a contradiction.

b) Since $2^2, 2^3, \ldots, 2^{\log_2 n}$ are all composite numbers that are less than n we have

$$
\begin{aligned}
COMP \ge & \quad (\log_2 n) - 1 \\
n - \pi(n) \ge & \quad (\log_2 n) - 1 \\
\pi(n) \le & \quad n + 1 - \log_2 n
\end{aligned}
$$

∎

EXERCISE: Improve the constants in the above theorem.

3 Definitions

Convention 3.1 Henceforth p will always denote a prime. Throughout this paper log means $\log_2$ unless otherwise noted.

Note that $\pi(n) = \sum_{p \le n} 1$. This is hard to prove things about. Hence we consider the following two functions.

Def 3.2

1. $f(n) = \sum_{p \le n} \log p$.
2. $g(n) = \sum_{k \ge 1} \sum_{p^k \le n} \log p$.

We will obtain upper and lower bounds on $\pi(n)$ in terms of $f(n)$. Then we will obtain upper and lower bounds on $f(n)$. The lower bound on $f(n)$ will use $g(n)$.

4 Bounds on $\pi(n)$ in terms of $f(n)$

Lemma 4.1 $\frac{f(n)}{\log n} \leq \pi(n)$.

Proof: $f(n) = \sum_{p \leq n} \log p \leq \pi(n) \log n$. Hence $\frac{f(n)}{\log n} \leq \pi(n)$. ▮

Lemma 4.2 *Let* $0 < \delta < 1$. *Then* $\pi(n) \leq \frac{f(n)}{\delta \log n} + n^\delta$.

Proof:

$$f(n) = \sum_{p \leq n} \log p \geq \sum_{n^\delta \leq p \leq n} \log p \geq (\pi(n) - \pi(n^\delta) \log n^\delta \geq (\pi(n) - n^\delta)\delta \log n.$$

So $\frac{f(n)}{\delta \log n} \geq \pi(n) - n^\delta$. Hence $\pi(n) \leq \frac{f(n)}{\delta \log n} + n^\delta$. ▮

Note 4.3 By Lemmas 4.1 and 4.2 we need only show that $f(n) = \Theta(n)$ to obtain the Weak Prime Number Theorem.

5 Upper Bound on $f(n)$

Lemma 5.1 $f(n) \leq 2n$.

Proof:
We obtain a recurrence for f.

$$\begin{aligned} f(2n) = \sum_{p \leq 2n} \log p = & \sum_{p \leq n} \log p + \sum_{n+1 \leq p \leq 2n} \log p \\ = & f(n) + \sum_{n+1 \leq p \leq 2n} \log p \\ = & f(n) + \log(\prod_{n+1 \leq p \leq 2n} p). \end{aligned}$$

We seek bounds on $\prod_{n+1 \leq p \leq 2n} p$. KEY IDEA: to bound a number find a number that it divides.

Clearly $\prod_{n+1 \leq p \leq 2n} p$ divides $(n+1)(n+2)\cdots 2n$. But this is large. We will divide $(n+1)(n+2)\cdots 2n$ by some quantity so that what we have left (a) is still an integer, and (b) still has $\prod_{n+1 \leq p \leq 2n} p$ dividing it.

Look at $\frac{(n+1)(n+2)\cdots 2n}{n!} = \binom{2n}{n}$. This is an integer. Since $\prod_{n \leq p \leq 2n} p$ divides the numerator but is relatively prime to the denominator, $\prod_{n+1 \leq p \leq 2n} p$ divides $\binom{2n}{n}$. Hence

$$\begin{aligned} \prod_{n+1 \leq p \leq 2n} & \leq \binom{2n}{n} \\ \prod_{n+1 \leq p \leq 2n} & \leq 2^{2n} \\ \log(\prod_{n+1 \leq p \leq 2n}) & \leq 2n \end{aligned}$$

Hence
$f(2n) \leq f(n) + 2n$.

Note that $f(2n-1) = f(2n)$ so we have
$f(2n-1) \le f(n) + 2n$.
These two equation together easily yields $f(n) \le 2n$.
$\blacksquare$

6 Lower Bounds on $f(n)$

To obtain lower bounds on $f(n)$ we first need to relate $g(n)$ to $f(n)$ and then get lower bounds on $g(n)$.

Lemma 6.1 $g(n) \le f(n) + 2\sqrt{n}\log n$.

Proof: $g(n) = \sum_{k\ge1}\sum_{p^k\le n}\log p$.

Let $1 \le p \le n$. How many times is $\log p$ a summand? Since $p^1 \le n$, at least once. If $p^2 \le n$ then it will be counted again. Hence, all primes $p \le n^{1/2}$ contribute at least two $\log p$ summands. More generally, if $p \le n^{1/i}$ then $\log p$ will appear i times as a summand. Hence we obtain
$g(n) = f(n) + f(n^{1/2}) + f(n^{1/3}) + \cdots + f(n^{1/\log n})$.

So $g(n) \le f(n) + (\log n)f(\sqrt{n})$. By Lemma 5.1 we get $g(n) \le f(n) + 2\sqrt{n}\log n$. $\blacksquare$

We now obtain a lower bound on $g(n)$.

Lemma 6.2 *For all $\epsilon > 0$ there exists n_0 such that $(\forall n \ge n_0)[g(n) \ge (1-\epsilon)n]$.*

Proof:
$g(2n) = \sum_{k\ge1}\sum_{p^k\le 2n}\log p$.

Fix p. How many times does $\log p$ appear as a summand? It will appear k times where $p^k \le 2n \le p^{k+1}$. This is $\lfloor \log_p 2n \rfloor$ times. Hence $g(2n) = \sum_{p\le 2n}(\lfloor \log_p 2n \rfloor)(\log p)$.

CLEVER IDEA- find some other quantity that is about the same.

Look at $\binom{2n}{n}$. All its prime factors are $\le 2n$.

Notation 6.3 If $p, m \in N$, $p \le m$, and p is prime then let $LARDIV_{p,m}$ be the largest i such that p^i divides m. Note that $LARDIV_{p,m!} = \left\lfloor \frac{m}{p} \right\rfloor + \left\lfloor \frac{m}{p^2} \right\rfloor + \left\lfloor \frac{m}{p^3} \right\rfloor + \cdots$.

Let $k_p = LARDIV_{p,(2n)!} - LARDIV_{p,n!n!} = LARDIV_{p,(2n)!} - 2LARDIV_{p,n!}$.
$\binom{2n}{n} = \prod_{p\le 2n} p^{k_p}$

We need to estimate k_p.

$$k_p = \sum_{i=1}^{\infty}\left\lfloor \frac{(2n)}{p^i} \right\rfloor - 2\sum_{i=1}^{\infty}\left\lfloor \frac{n}{p^i} \right\rfloor = \sum_{i=1}^{\infty}\left\lfloor \frac{2n}{p^i} \right\rfloor - 2\left\lfloor \frac{n}{p^i} \right\rfloor$$

Note that each summand is either 0 or 1. Also note that at most $\lfloor \log_p(2n) \rfloor$ of the terms are nonzero. Hence $k_p \le \lfloor \log_p(2n) \rfloor$.

So

$$\binom{2n}{n} = \prod_{p \le 2n} p^{k_p} \le \prod_{p \le 2n} p^{\lfloor \log_p 2n \rfloor}$$

By Stirling's formula $\binom{2n}{n} = \Theta(\frac{2^{2n}}{\sqrt{n}})$. Let n_0' be such that, for all $n \ge n_0$, $\binom{2n}{n} \ge \frac{2^{2n}}{n}$. Hence we have

$$\frac{2^{2n}}{n} \le \prod_{p \le 2n} p^{\lfloor \log_p(2n) \rfloor}.$$

$$2n - \log n \le \sum_{p \le 2n} (\lfloor \log_p(2n) \rfloor)(\log p) = g(2n).$$

$$g(2n) \ge 2n - \log n.$$

Let n_0'' be the least number bigger than n_0' such that $(\forall n \ge n_0)[g(2n) \ge (2-\epsilon)n]$. Since $g(2n-1) = g(2n)$, we also have $g(2n-1) \ge (2-\epsilon)n$. Let $n_0 = 2n_0''$. We have $(\forall n \ge n_0)[g(n) \ge (2-\epsilon)n]$. ∎

Lemma 6.4 *For all $\epsilon > 0$ there exists n_0 such that $(\forall n \ge n_0)[f(n) \ge (1-\epsilon)n]$.*

Proof: By Lemma 6.1 we have $f(n) \ge g(n) - 2\sqrt{n}\log n$. Let $0 < \epsilon' < \epsilon$. By Lemma 6.2 we have that there exists n_0' such that

$(\forall n \ge n_0)[g(n) \ge (2-\epsilon')n$

Hence we have

$$f(n) \ge (2-\epsilon')n - 2\sqrt{n}\log n.$$

Let n_0 be the least number $\ge n_0'$ such that

$$(\forall n \ge n_0)[(2-\epsilon')n - 2\sqrt{n}\log n \ge (2-\epsilon)n].$$

Clearly $(\forall n \ge n_0)[f(n) \ge (2-\epsilon)n]$. ∎

7 The Weak Prime Number Theorem

Theorem 7.1 *For all ϵ there exists n_0 such that*

$$(\forall n \ge n_0)[(1-\epsilon)\frac{n}{\log n} \le \pi(n) \le (2+\epsilon)\frac{n}{\log n}]$$

Proof: By Lemma 4.1 we have that $\frac{f(n)}{\log n} \le \pi(n)$. By Lemma 6.4 we have

$$(\exists n_0')(\forall n \ge n_0')[f(n) \ge (1-\epsilon)n].$$

Hence we have

$$(\exists n_0')(\forall n \ge n_0')[(1-\epsilon)\frac{n}{\log n} \le \pi(n)].$$

By Lemma 4.2 we have that, for any δ with $0 < \delta < 1$, $\pi(n) \le \frac{f(n)}{\delta \log n} + n^{\delta}$. By Lemma 5.1 we have $f(n) \le 2n$. Hence we have

$$\pi(n) \le \frac{2n}{\delta \log n} + n^{\delta} = \frac{2}{\delta}\frac{n}{\log n} + n^{\delta}$$

Let $\delta > 0$ be such that $\frac{2}{\delta} < (2+\epsilon)$. Let n_0'' be such that

$$(\forall n \ge n_0'')[\frac{2}{\delta}\frac{n}{\log n} + n^{\delta} \le (2+\epsilon)\frac{n}{\log n}].$$

Let $n_0 = \max\{n_0', n_0''\}$. Then, for all $n \ge n_0$,

$$(1-\epsilon)\frac{n}{\log n} \le \pi(n) \le (2+\epsilon)\frac{n}{\log n}.$$

▮

Note 7.2 The real Prime Number Theorem uses $\ln n$ instead of $\log n$. To see how the weak one compares, we rewrite it using the fact that $\log n = \frac{\ln n}{\ln 2}$. We have

$$(\forall \epsilon > 0)(\exists n_0)(\forall n \ge n_0)[(1-\epsilon)(\ln 2)\frac{n}{\ln n} \le \pi(n) \le (2+\epsilon)(\ln 2)\frac{n}{\ln n}].$$

Using $0.69 \le \ln 2 \le 0.7$.

$$(\forall \epsilon > 0)(\exists n_0)(\forall n \ge n_0)[0.69\frac{n}{\ln n} \le \pi(n) \le 1.4\frac{n}{\ln n}.]$$

Note 7.3 Chebyshev obtained

$$(\forall \epsilon > 0)(\exists n_0)(\forall n \ge n_0)[0.875\frac{n}{\ln n} \le \pi(n) \le 1.125\frac{n}{\ln n}.$$

EXERCISE: For $\epsilon = 1, \frac{1}{2}, \ldots$ find a value of n_0. Try to make it as small as possible.

We can now prove a weak version of Bertrand's Postulate.

Theorem 7.4 *There exists n_0 such that, for all $n \ge n_0$, there is a prime between n and $3n$.*

Proof:

We need to show that $\pi(3n) - \pi(n) \geq 1$.

By Theorem 7.1 with $\epsilon = \frac{1}{8}$ we have $(\exists n_0)(\forall n \geq n_0)[(\frac{7}{8})\frac{n}{\log n} \leq \pi(n) \leq \frac{17}{8}\frac{n}{\log n}]$. Hence

$\pi(3n) \geq \frac{21}{8}\frac{n}{\log 3n} \geq \frac{21}{8}\frac{n}{\log n}$. and $\pi(n) \leq \frac{17}{8}\frac{n}{\log n}$.

Hence $\pi(3n) - \pi(n) \geq 1$. ∎

EXERCISE: Prove a tighter version of Bertrand's postulate using these methods.

EXERCISE: Using cruder approximations that work for all n, obtain a weaker version of the Weak Prime Number Theorem that works for all n. (It will not have an ϵ or n_0 in it.)