Proving Infinitude of Prime Numbers Using Binomial Coefficients

Phuong Nguyen
University of Toronto

June 10, 2008

Abstract

We study the problem of proving in weak theories of Bounded Arithmetic the theorem that there are arbitrarily large prime numbers. We show that the theorem can be proved by some "minimal" reasoning (i.e., in the theory $\mathbf{I\Delta}_0$) using concepts such as (the logarithm) of a binomial coefficient. In fact we prove Bertrand's Postulate (that there is at least a prime number between n and $2n$, for all $n > 1$) and the fact that the number of prime numbers between n and $2n$ is of order $\Theta(n/\ln(n))$. The proofs that we formalize are much simpler than several existing formalizations, and our theory turns out to be a sub-theory of a recent theory proposed by Woods and Cornaros that extends $\mathbf{I\Delta}_0$ by a special counting function.

1 Introduction

A long standing problem in proof complexity theory is whether the fact that there are infinitely many prime numbers is provable in the theory $\mathbf{I\Delta}_0$, the theory over the vocabulary $0, 1, +, \cdot, <$ that is axiomatized by basic properties of this vocabulary and induction axioms for all bounded formulas. The problem remains open even when we replace $\mathbf{I\Delta}_0$ by $\mathbf{I\Delta}_0(\pi)$, a theory that extends $\mathbf{I\Delta}_0$ by adding the function $\pi(n)$ which is the number of prime numbers less than or equal to n [Woo81]. ($\mathbf{I\Delta}_0(\pi)$ is also called $\mathbf{I\Delta}_0(\pi) + \mathit{def}(\pi)$ in the literature.) The motivation for the latter is: suppose that we are able to count the number of primes, then is it possible to prove the infinitude of primes using some "minimal" reasoning?

These problems belong to the area recently named Bounded Reverse Mathematics [Coo07] whose purpose is to formalize and prove (the discrete versions of) mathematical theorems in weak theories of Bounded Arithmetic. A related problem [PWW88] is whether a weak form of the Pigeonhole Principle is provable in $\mathbf{I\Delta}_0$, or equivalently, whether it has polynomial-size constant-depth Frege proofs.

Recently some progress has been made in [WC07] where it is shown that $\mathbf{I\Delta}_0(\xi)$ (called $\mathbf{I\Delta}_0(\xi)+\mathit{def}(\xi)$ in [WC07]) proves the infinitude of primes. Here $\mathbf{I\Delta}_0(\xi)$ extends $\mathbf{I\Delta}_0$ by the function ξ that counts some definable sets of prime numbers. The function π can be defined using ξ, so $\mathbf{I\Delta}_0(\xi)$ is an extension of $\mathbf{I\Delta}_0(\pi)$. It is unlikely that ξ can be defined in $\mathbf{I\Delta}_0(\pi)$.

In an earlier paper [Cor95] it is shown that the infinitude of primes is also provable in $\mathbf{I\Delta}_0(\pi, K)$, the theory that extends $\mathbf{I\Delta}_0(\pi)$ by a defining axiom for the function

$$K(n) = \sum_{i=1}^{n} \ln(i)$$

It is not clear whether $\mathbf{I\Delta}_0(\xi)$ extends $\mathbf{I\Delta}_0(\pi, K)$, or vice versa.

In this paper we show that the infinitude of prime numbers is provable in $\mathbf{I\Delta}_0(\pi, \mathit{lbc})$, the theory obtained from $\mathbf{I\Delta}_0(\pi)$ by adding a defining axiom for the function

$$\mathit{lbc}(n) = \ln(\frac{(2n)!}{n!n!})$$

(*lbc* stands for *logarithm of binomial coefficient*). We also show that the function *lbc* is definable in $\mathbf{I\Delta}_0(\xi)$. Together with the fact proved in [WC07] that π is definable in $\mathbf{I\Delta}_0(\xi)$, this implies that $\mathbf{I\Delta}_0(\pi, \mathit{lbc})$ is a sub-theory of $\mathbf{I\Delta}_0(\xi)$. So our results strengthen the results from [WC07]. On the other hand, we do not know whether our theory extends that of [Cor95], or vice versa.

Note that the function ξ [WC07] is a counting function that is more general than π, while both K [Cor95] and our function *lbc* are not. Also, if we add to $\mathbf{I\Delta}_0$ a counting function and its defining axiom for every $\mathbf{\Delta}_0$-definable set, then the resulting theory, here we called $\mathbf{I\Delta}_0(\mathit{count})$, extends all $\mathbf{I\Delta}_0(\xi)$, $\mathbf{I\Delta}_0(\pi, K)$ and $\mathbf{I\Delta}_0(\pi, \mathit{lbc})$. It has been shown [CD94] that $\mathbf{I\Delta}_0(\mathit{count})$ proves the Prime Number Theorem (that there are $\Theta(n/\ln(n))$ primes less than n). It is easy to see that $\mathbf{I\Delta}_0(\mathit{count})$ is equivalent to the number part of the theory $\mathbf{VTC}^0$ [NC05, CN06], a two-sorted theory that is associated with the two-sorted complexity class $\mathbf{TC}^0$.

1.1 Existing Formalizations

Our formalization is based on [Ngu08, Chapter 8]. At high level, the proof that we choose to formalize is essentially the same as that of [WC07]. However, we explicitly use the binomial coefficients mentioned above, so our formalization is simpler. In fact, the axiom that we need to define *lbc* is provable (in $\mathbf{I\Delta}_0$) from the defining axiom for the function ξ introduced in [WC07]. Moreover, the function ξ seems to be indispensable for the formalization in [WC07], because it is needed in proving (the approximate version of) the asymptotic identity

$$(\psi(x) - \psi(\frac{x}{2}) + \psi(\frac{x}{3}) - \psi(\frac{x}{4}) + \ldots) = x\ln(2)$$

where

$$\psi(x) = \sum_{i \leq x} \Lambda(i)$$

and $\Lambda(x)$ is the von Mangoldt function,

$$\Lambda(x) = \begin{cases} \ln(p) & \text{if } x = p^j \text{ for some prime } p \text{ and some } j \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

1.2 Our Formalizations

The proofs that we formalized are simple proofs which rely on different (approximate) representations of

$$\ln(\frac{(2n)!}{n!n!}) \tag{1}$$

One way of computing (1) is to use the fact that

$$\sum_{i=1}^{n} \ln(i) = n\ln(n) - n + \mathcal{O}(\ln(n)) \tag{2}$$

This produces

$$\ln(\frac{(2n)!}{n!n!}) = \sum_{i=1}^{2n} \ln(i) - 2\sum_{i=1}^{n} \ln(i) = 2n\ln(2) + \mathcal{O}(\ln(n)) \tag{3}$$

Another expression for (1) is

$$\sum_{p \leq 2n} \left(\ln(p) \sum_{1 \leq j \wedge p^j \leq 2n} (\lfloor 2n/p^j \rfloor - 2\lfloor n/p^j \rfloor) \right) \tag{4}$$

This expression reveals useful information about the prime numbers that are $\leq 2n$. For example, it gives us

$$\ln(\frac{(2n)!}{n!n!}) \leq \pi(2n)\ln(2n)$$

and so a lower bound for $\pi(2n)$ follows using (3). Moser's simple proof of Bertrand's Postulate that we formalize also stems from (4) (see Lemma 4.2).

In our formalizations, the function *lbc* is defined based on the expression (4). The obstacle that prevents us from resolving Woods' conjecture is the inability to compute in $\mathbf{I\Delta}_0(\pi)$ this summation.

Of course we cannot compute the function $\ln(x)$ precisely, so as in [Woo81] we use an approximation to it. Our approximation and much of the formalizations are from [Ngu08, Chapter 8]. The approximation to $\ln(x)$, denoted by $\ln(x,m)$ for a parameter m, is essentially the same as the approximation given in [Woo81]. Here we give a more detailed and direct proof of our version of (2).

1.3 Organization

The paper is organized as follows. In Section 2 we recall $\mathbf{I\Delta_0}$ and some important properties. In Section 2.2 we define in $\mathbf{I\Delta_0}$ an approximation to $\ln(x)$. The function *lbc* is defined in Section 2.7, and in Section 2.8 we show that it is definable in $\mathbf{I\Delta}_0(\xi)$. The $\mathbf{I\Delta}_0(\pi, lbc)$-proof of a lower bound for $\pi(n)$ is given in Section 3. The lower bound for $\pi(2n) - \pi(n)$ and Bertrand's Postulate are proved in Section 4.

2 The Theories $\mathbf{I\Delta}_0$, $\mathbf{I\Delta}_0(\pi)$, and $\mathbf{I\Delta}_0(\pi, lbc)$

The language of $\mathbf{I\Delta_0}$ is

$$\{0, 1, +, \cdot, <, =\}$$

The theory $\mathbf{I\Delta_0}$ is axiomatized by some basic defining axioms for the symbols in the language (see [HP93, Kra95, CN06]) and induction axiom scheme for bounded formulas. $\overline{\mathbf{I\Delta}}_0$ denotes the universal conservative extension of $\mathbf{I\Delta_0}$ obtained by adding Skolem functions that eliminate quantifiers in the axioms of $\mathbf{I\Delta}_0$. (We do not need the fact that $\overline{\mathbf{I\Delta}}_0$ is a universal theory here.)

(Instead of $\mathbf{I\Delta_0}$ and its extensions, we can use the two-sorted theory $\mathbf{V}^0$ [CN06] and its corresponding extensions, because $\mathbf{V}^0$ is conservative over $\mathbf{I\Delta_0}$ and the same can be shown for their respective extensions. Care should be taken, however, when we look at the associated complexity classes: $\mathbf{V}^0$ is associated with the two-sorted class $\mathbf{AC}^0$ where sets are presented by binary strings and numbers by unary strings; on the other hand, $\mathbf{I\Delta_0}$ is associated with the Linear Time Hierarchy, because here numbers are written in binary.)

The following theorem is from [Ben62, HP93, Bus98, CN06]:

Theorem 2.1. *The relation (on numbers) $y = z^x$ can be represented by a $\mathbf{\Delta}_0$ formula.*

Corollary 2.2. *The function $|x|$ (or also $\log(x)$), where $|0| = 0$ and $|x| = \lfloor \log_2(x) \rfloor$ if $x \geq 1$, is definable in $\mathbf{I\Delta}_0$.*

The following theorem is from [Woo81]:

Theorem 2.3. *For a bounded $\mathbf{\Delta}_0$-sequence $x_1, x_2, \ldots, x_\ell$ where $\ell \leq (\log(a))^d$ for some a and some constant $d \in \mathbb{N}$, the function*

$$\sum_{1 \leq i \leq \ell} x_i$$

is definable in $\mathbf{I\Delta}_0$ and it is provable in $\overline{\mathbf{I\Delta}}_0$ that

$$\sum_{1 \leq i \leq \ell+1} x_i = \sum_{1 \leq i \leq \ell} x_i + x_{\ell+1}$$

2.1 Rational Numbers in $\mathbf{I\Delta}_0$

We will approximate the natural logarithm function by rational numbers. Here we only need nonnegative numbers which can be defined in $\mathbf{I\Delta}_0$ by pairs $\langle x, y\rangle$, where

$$\langle x, y\rangle =_{\text{def}} (x+y)(x+y+1)+2y$$

For readability we will write $\frac{x}{y}$ for $\langle x, y\rangle$. Equality, inequality, addition and multiplication for rational numbers are defined in the standard way, and these are preserved under the embedding $x \mapsto \frac{x}{1}$. For example, $=_{\mathbb{Q}}$ and $\leq_{\mathbb{Q}}$ are defined as:

$$\frac{x}{y} =_{\mathbb{Q}} \frac{x'}{y'} \equiv xy' = x'y, \qquad \text{and} \qquad \frac{x}{y} \leq_{\mathbb{Q}} \frac{x'}{y'} \equiv xy' \leq x'y$$

Then it can be shown that

$$\mathbf{I\Delta}_0 \vdash \lfloor x/y \rfloor \leq_{\mathbb{Q}} \frac{x}{y} <_{\mathbb{Q}} \lfloor x/y \rfloor + 1$$

(here $\lfloor x/y \rfloor = max\{z \ : \ zy \leq x\}$, and $r <_{\mathbb{Q}} s \equiv (r \leq_{\mathbb{Q}} s \wedge r \neq_{\mathbb{Q}} s)$). In the following discussion, we will simply omit the subscript $\mathbb{Q}$ from $=_{\mathbb{Q}}$, $\leq_{\mathbb{Q}}$, etc.; the exact meaning will be clear from the context.

For a rational number $\frac{r}{s} \geq 1$, define

$$|\frac{r}{s}| = max\{i \ : \ s2^i \leq r\}$$

2.2 Approximating $\ln(x)$ in $\mathbf{I\Delta}_0$

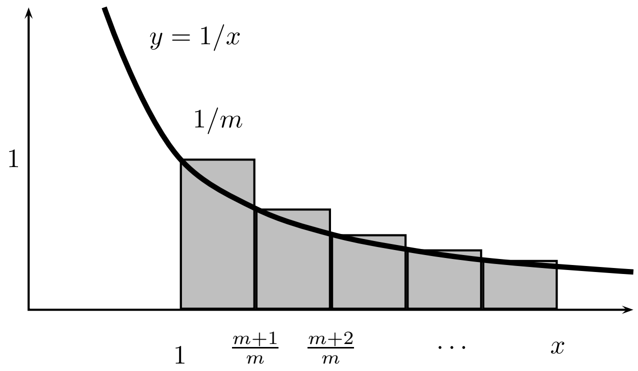

Figure 1: Defining $\ln(x, m)$ for $1 \leq x \leq 2$: the shaded area is (6).

We will now define in $\mathbf{I\Delta}_0$ a function $\ln(x, m)$ which approximates $\ln(x)$ up to $\mathcal{O}(|x|/m)$, for $x \in \mathbb{N}$, where m is a polynomial in $|a|$. Following [Woo81] we will first define $\ln(x, m)$ that approximates $\ln(x)$ upto $1/m$ for $1 \leq x \leq 2$. Then for $x > 2$ define

$$\ln(x, m) = |x| \ln(2, m) + \ln(\frac{x}{2^{|x|}}) \tag{5}$$

It is easy to see that for any $x > 1$, $\ln(x, m)$ approximates $\ln(x)$ upto $\mathcal{O}(|x|/m)$.

Our definition of $\ln(x, m)$ for $1 \le x \le 2$ is essentially the same as the definition of $\log^+$ of [Woo81]. Note that

$$\ln(x) = \int_1^x \frac{1}{y} dy$$

Our approximation will be roughly (the shaded area in Figure 1):

$$\sum_{m \le k < \lceil mx \rceil} \frac{1}{m} \frac{1}{k/m} = \sum_{m \le k < \lceil mx \rceil} \frac{1}{k} \tag{6}$$

We will not compute this summation precisely (since we want to avoid computing the common denominator). Instead we approximate $\frac{1}{k}$ by $\frac{\lfloor b/k \rfloor}{b}$ for some b determined below. Thus

$$\ln(x, m) = \frac{\sum_{m \le k < \lceil mx \rceil} \lfloor b/k \rfloor}{b} \tag{7}$$

The summation in (7) can be carried out in $\mathbf{I\Delta}_0$ by Theorem 2.3.

Notice that (6) is an upper bound for $\ln(x)$ with an error (the total area of the shaded region above the line $xy = 1$) at most $1/m$, and (7) is a lower bound for (6) with an error at most $\lceil mx \rceil / b$. So to get an $1/m$-approximation to $\ln(x)$ it suffices to take $b = m^3$.

Notation Throughout this paper, fix some a sufficiently large and m a power of 2, $m = \mathit{polylog}(a) = 2^h$. (In particular, $m > |a|^2$.) We use $\|\cdot\|$ for absolute value, e.g., $\|t_1 - t_2\| \le s$ is an abbreviation for $t_1 \le t_2 + s \wedge t_2 \le t_1 + s$.

Definition 2.4 ($\ln(x, m)$ or just $\ln(x)$)**.** *For* $1 \le x \le 2$*,* $\ln(x, m)$ *is defined as in* (7) *with* $b = m^3$*. For* $x > 2$*,* $\ln(x, m)$ *is defined as in* (5)*.*

Lemma 2.5 (Provable in $\overline{\mathbf{I\Delta}}_0$)**. a)** $x \le y \supset \ln(x, m) \le \ln(y, m)$.
b) $\|\ln(xy, m) - (\ln(x, m) + \ln(y, m))\| = \mathcal{O}(\frac{|x|+|y|}{m})$

Proof. Part **a)** is straightforward from definition. For part **b)** we consider the following cases.

(i) $1 \le x, y \le 2$ and $xy \le 2$. By definition we have

$$\ln(xy, m) = \frac{1}{b} \sum_{m \le k < \lceil mxy \rceil} \lfloor b/k \rfloor = \ln(x, m) + \frac{1}{b} \sum_{\lceil mx \rceil \le k < \lceil mxy \rceil} \lfloor b/k \rfloor$$

Hence

$$\ln(xy, m) - \ln(x, m) - \ln(y, m) = \frac{1}{b} \left(\sum_{\lceil mx \rceil \le k < \lceil mxy \rceil} \lfloor b/k \rfloor - \sum_{m \le k < \lceil my \rceil} \lfloor b/k \rfloor \right)$$

Let r, t be such that

$$\frac{r-1}{m} < x \leq \frac{r}{m}, \qquad \frac{t-1}{m} < y \leq \frac{t}{m}$$

Then $m < r, t \leq 2m$, and

$$\frac{(r-1)(t-1)}{m} < \lceil mxy \rceil \leq \frac{rt}{m}$$

Now

$$\begin{aligned}
\sum_{\lceil mx \rceil \leq k < \lceil mxy \rceil} \lfloor b/k \rfloor &\leq \sum_{r \leq k < \lceil rt/m \rceil} \lfloor b/k \rfloor \\
&< \sum_{r \leq k < \lceil rt/m \rceil} (m \lfloor b/km \rfloor + m) \\
&\leq \sum_{r \leq k < \lceil rt/m \rceil} \left(m + \sum_{(k-1)m \leq i < km} \lfloor b/i \rfloor \right) \\
&< rt + \sum_{(r-1)m \leq i < rt} \lfloor b/i \rfloor
\end{aligned}$$

Also,

$$\begin{aligned}
\sum_{\lceil mx \rceil \leq k < \lceil mxy \rceil} \lfloor b/k \rfloor &\geq \sum_{r \leq k < \lceil (r-1)(t-1)/m \rceil} \lfloor b/k \rfloor \\
&\geq \sum_{r \leq k < \lceil (r-1)(t-1)/m \rceil} m \lfloor b/km \rfloor \\
&\geq \sum_{r \leq k < \lceil (r-1)(t-1)/m \rceil} \left(\sum_{km \leq i < (k+1)m} \lfloor b/i \rfloor \right) \\
&= \sum_{rm \leq i < (r-1)(t-1)} \lfloor b/i \rfloor
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sum_{m \leq k < my} \lfloor b/k \rfloor &= \sum_{m \leq k < t} \lfloor b/k \rfloor \\
&< \sum_{m \leq k < t} (r \lfloor b/kr \rfloor + r) \\
&\leq \sum_{m \leq k < t} \left(r + \sum_{(k-1)r \leq i < kr} \lfloor b/i \rfloor \right) \\
&< rt + \sum_{(m-1)r \leq i < rt} \lfloor b/i \rfloor
\end{aligned}$$

And,

$$\begin{aligned}\sum_{m\le k<my} \lfloor b/k\rfloor &= \sum_{m\le k<t} \lfloor b/k\rfloor \\ &\ge \sum_{m\le k<t} r\lfloor b/kr\rfloor \\ &\ge \sum_{m\le k<t} \left(\sum_{kr\le i<(k+1)r} \lfloor b/i\rfloor\right) \\ &= \sum_{rm\le i<rt} \lfloor b/i\rfloor\end{aligned}$$

As a result, we can derive an upper bound for

$$\|\ln(xy,m) - (\ln(x,m) + \ln(y,m))\|$$

by noting that $m < r, t \le 2m$ and $m = \mathit{polylog}(a)$ for some a sufficiently large.

(ii) $1 \le x, y \le 2$ and $2 < xy \le 4$. First, by the same argument we can show that

$$\ln(z,m) = \sum_{m\le k<mz} \lfloor b/k\rfloor + \mathcal{O}(\frac{1}{m})$$

for $2 < z \le 4$. The current case is handled using this and the same arguments as in the previous case.

(iii) Finally, the case where $x, y > 2$ is reduced to the previous cases using (5). □

2.3 Defining $\sum \ln(i)$ in $\mathbf{I\Delta}_0$

The fact that $\sum_{1\le i\le n} \ln(i)$ is definable in $\mathbf{I\Delta}_0$ is from [Woo81]. We reprove it here (for our definition of $\ln(x)$) in order to roughly estimate the sum.

Theorem 2.6. a) *The following function is definable in* $\mathbf{I\Delta}_0$*:*

$$\sum_{i=1}^{n} \ln(i) \tag{8}$$

b) *Let*

$$S = \sum_{i=1}^{m} \ln(i), \qquad T = \sum_{t=1}^{m} \ln(\frac{m+t}{m}), \qquad T_n = \sum_{i=2^{|n-1|}+1}^{n} \ln(\frac{i}{2^{|n-1|}}) \tag{9}$$

Then S, T, T_n *are definable in* $\mathbf{I\Delta}_0$*, and it is provable in* $\overline{\mathbf{I\Delta}_0}$ *that (let* $\ell = |n-1|$*)*

$$\sum_{i=1}^{n} \ln(i) = S + (n\ell - 2^{\ell+1} - (h-2)2^h)\ln(2) + (2^{\ell-h} - 1)T + T_n \tag{10}$$

c) *It is provable in* $\overline{\mathbf{I\Delta}_0}$ *that*

$$\sum_{i=1}^{n+1} \ln(i) = \sum_{i=1}^{n} \ln(i) + \ln(n+1) \tag{11}$$

Proof. **a)** and **b)** First, S and T are definable in $\mathbf{I\Delta}_0$ by Theorem 2.3 because the summations have length $m = \mathit{polylog}(a)$. The fact that T_n is also definable in $\mathbf{I\Delta}_0$ will be shown in the following discussion.

Suppose that $n > m$. Recall that $\ell = |n-1|$, i.e.,

$$2^\ell < n \leq 2^{\ell+1}$$

To compute (8), we first compute the following sums (recall $m = 2^h$):

$$S_j = \sum_{i=2^j+1}^{2^{j+1}} \ln(i) \quad \text{for } h \leq j < \ell, \qquad S_\ell = \sum_{i=2^\ell+1}^{n} \ln(i) \tag{12}$$

Then (8) is $S + \sum_{j=h}^{\ell} S_j$, and therefore can be computed in $\mathbf{I\Delta}_0$ by Theorem 2.3.

To compute S_j, note that for $2^j < i \leq 2^{j+1}$, by definition we have

$$\ln(i) = j\ln(2) + \ln(\frac{i}{2^j})$$

Hence

$$S_j = \sum_{i=2^j+1}^{2^{j+1}} \ln(i) = j2^j \ln(2) + \sum_{i=2^j+1}^{2^{j+1}} \ln(\frac{i}{2^j})$$

To compute

$$\sum_{i=2^j+1}^{2^{j+1}} \ln(\frac{i}{2^j})$$

notice that

$$\ln(\frac{i}{2^j}) = \sum_{s=m}^{m+t-1} \lfloor b/s \rfloor = \ln(\frac{m+t}{m}) \qquad \text{for } 2^j + (t-1)2^{j-h} + 1 \leq i \leq 2^j + t2^{j-h} \tag{13}$$

So

$$\sum_{i=2^j+1}^{2^{j+1}} \ln(\frac{i}{2^j}) = 2^{j-h} \sum_{t=1}^{m} \ln(\frac{m+t}{m}) = 2^{j-h} T$$

Therefore

$$S_j = j2^j \ln(2) + 2^{j-h} T$$

Similarly,

$$S_\ell = (n - 2^\ell)\ell \ln(2) + \sum_{i=2^\ell+1}^{n} \ln(\frac{i}{2^\ell}) = (n-2^\ell)\ell\ln(2) + T_n$$

Note that by (13),

$$T_n = 2^{\ell-h}\sum_{t=1}^{t_n}\ln(\frac{m+t}{m}) + (n-2^\ell - t_n 2^{\ell-h})\ln(\frac{m+t_n+1}{m})$$

where $t_n = \lfloor (n-1-2^\ell)/2^{\ell-h}\rfloor$, i.e.,

$$2^\ell + t_n 2^{\ell-h} + 1 \le n \le 2^\ell + (t_n+1)2^{\ell-h}$$

This shows that T_n is definable in $\mathbf{I\Delta}_0$, because $t_n \le m$. It follows, in addition, that

$$0 < T_n \le 2^{\ell-h}T$$

Also, it is easy to see that for $n > m$:

$$T_{2n} = 2T_n$$

As a result, (8) is

$$\begin{aligned} S+\sum_{j=h}^{\ell} S_j &= S + \left(\sum_{j=h}^{\ell-1} j2^j\ln(2) + 2^{j-h}T\right) + (n-2^\ell)\ell\ln(2) + T_n \\ &= S + (\sum_{j=h}^{\ell-1} j2^j)\ln(2) + (n-2^\ell)\ell\ln(2) + (2^{\ell-h}-1)T + T_n \\ &= S + ((\ell-2)2^\ell - (h-2)2^h)\ln(2) + (n-2^\ell)\ell\ln(2) + (2^{\ell-h}-1)T + T_n \end{aligned}$$

The last equality follows from the fact (provable in $\overline{\mathbf{I\Delta}}_0$ by induction on i) that

$$\sum_{j=1}^{i} j2^j = (i-1)2^{i+1} + 2$$

From the last equation we can derive (10).

c) The fact that (11) are provable in $\overline{\mathbf{I\Delta}}_0$ is straightforward from the above definition. □

2.4 $\mathbf{I\Delta}_0(\pi)$ and Defining $\sum \ln(p)$ in $\mathbf{I\Delta}_0(\pi)$

Notation Throughout this paper, the index p is used for prime numbers. $\mathcal{P}$ denotes the set of prime numbers. Note that the relation $x \in \mathcal{P}$ is represented by a $\mathbf{\Delta}_0$ formula.

Let

$$\pi(n) = \#\{p \le n \ : \ p \in \mathcal{P}\}$$

$\mathbf{I\Delta}_0(\pi)$ extends $\mathbf{I\Delta}_0$ by π and the following defining axioms for it:

$$\pi(0) = 0$$

$$\pi(n+1) = \begin{cases} \pi(n) & \text{if } n+1 \notin \mathcal{P} \\ \pi(n)+1 & \text{otherwise} \end{cases}$$

Chebyshev's function

$$\vartheta(x) = \sum_{p \le x} \ln(p) \tag{14}$$

plays an important role. Here we use

$$\vartheta(x, m) = \sum_{p \le x} \ln(p, m) \tag{15}$$

and will simply write $\vartheta(x)$ for $\vartheta(x, m)$. We use the following defining axioms for ϑ:

$$\vartheta(1) = 0, \qquad \vartheta(n+1) = \begin{cases} \vartheta(n) + \ln(n+1) & \text{if } n+1 \in \mathcal{P} \\ \vartheta(n) & \text{otherwise} \end{cases} \tag{16}$$

Theorem 2.7. *The function $\vartheta(x)$ with defining axioms* (16) *is definable in* $\mathbf{I\Delta}_0(\pi)$.

Proof Sketch. The proof is similar to the proof of Theorem 2.6. For example,

$$\sum_{2^j < p \le 2^{j+1}} \ln(p) = j(\pi(2^{j+1}) - \pi(2^j)) \ln(2) + \sum_{2^j < p \le 2^{j+1}} \ln(\frac{p}{2^j})$$

Using (13) we have

$$\sum_{2^j < p \le 2^{j+1}} \ln(\frac{p}{2^j}) = \sum_{t=1}^{m} (\pi(2^j + t2^{j-h}) - \pi(2^j + (t-1)2^{j-h})) \ln(\frac{m+t}{m})$$

Hence by Theorem 2.3 the sum

$$\sum_{2^j < p \le 2^{j+1}} \ln(p)$$

is definable in $\mathbf{I\Delta_0}(\pi)$, for $0 \le j \le |n|$.

Similarly, we can define

$$\sum_{2^\ell < p \le n} \ln(p)$$

where $\ell = |n-1|$. Therefore we can define

$$\sum_{p \le n} \ln(p) \qquad \text{and hence} \qquad \sum_{n \le p \le k} \ln(p)$$

in $\mathbf{I\Delta_0}(\pi)$.

The fact that (16) is provable in $\overline{\mathbf{I\Delta}}_0(\pi)$ is clear from the above definition.
□

2.5 Approximating $\ln(2)$

Lemma 2.8 (Provable in $\overline{\mathbf{I\Delta}}_0$)**.**

$$\|\ln(2,m) - \ln(2,2m)\| < \frac{1}{2m}$$

Proof. From definition we have

$$\begin{aligned}
\ln(2,2m) - \ln(2,m) &= \frac{1}{8m^3}\sum_{k=2m}^{4m-1}\lfloor 8m^3/k\rfloor - \frac{1}{m^3}\sum_{k=m}^{2m-1}\lfloor m^3/k\rfloor \\
&= \frac{1}{8m^3}\sum_{k=m}^{2m-1}(\lfloor 8m^3/2k\rfloor + \lfloor 8m^3/(2k+1)\rfloor - 8\lfloor m^3/k\rfloor)
\end{aligned}$$

For $m \le k < 2m$, let $\lfloor m^3/k\rfloor = q$, then it can be shown that

$$\begin{aligned}
4q \le \lfloor 8m^3/2k\rfloor &\le 4q+3 \\
4q - 2m \le \lfloor 8m^3/(2k+1)\rfloor &\le 4q+3
\end{aligned}$$

In other words, for $m \ge 3$ we have

$$\|\lfloor 8m^3/2k\rfloor + \lfloor 8m^3/(2k+1)\rfloor - 8\lfloor m^3/k\rfloor\| \le 2m$$

Consequently,

$$\|\ln(2,2m) - \ln(2,m)\| \le \frac{1}{8m^3}(2m-m)2m = \frac{1}{2m} \qquad \square$$

The lemma can be used to show that for any (standard) error $\epsilon \in \mathbb{Q}$, there is $m_0 \in \mathbb{N}$ so that

$$\|\ln(2,m) - \ln(2)\| < \epsilon$$

for all $m > m_0$, m is a power of 2. For example, from the lemma we have

$$\|\ln(2,2^h) - \ln(2,2^{k+h})\| < \frac{1}{2^h}$$

for all $k \ge 0$. Since $\ln(2,8) = \frac{368}{512} = \frac{23}{32}$, it follows that

$$\frac{19}{32} < \ln(2,2^h) < \frac{27}{32}$$

for $h \ge 3$.

2.6 Unique Prime Factorization

The Fundamental Theorem of Arithmetic (or Unique Prime Factorization Theorem) states that any natural number $n > 1$ can be written uniquely as

$$n = p_1^{e_1} \cdot p_2^{e_2} \cdot \ldots \cdot p_k^{e_k}$$

where $p_1 < p_2 < \ldots < p_k$ are prime numbers, and $e_i \geq 1$.

In $\mathbf{I\Delta}_0$ we can prove the existence and uniqueness of the sequence

$$(p_1, e_1), (p_2, e_2), \ldots, (p_k, e_k)$$

that contains all prime divisors of n, and $e_i \geq 1, p_i^{e_i} \mid n, p_i^{e_i+1} \nmid n$. Note that the sequence can be encoded by a binary string of length $\mathcal{O}(|n|)$. Also, the product

$$\prod_{i=1}^{k} p_i^{e_i}$$

for such sequence can be defined and proved to be n in $\mathbf{I\Delta}_0$.

Here we use the following function which is provably total in $\mathbf{I\Delta}_0$ (**ex** stands for exponent):

$$\mathbf{ex}(p, n) = max\{j \ : \ p^j | n\} \tag{17}$$

Our version of the Fundamental Theorem of Arithmetic is as follows:

Lemma 2.9. *The sum*

$$\sum_{p|n} \mathbf{ex}(p, n) \ln(p, m)$$

is definable in $\mathbf{I\Delta}_0$*, and it is provable in* $\overline{\mathbf{I\Delta}}_0$ *that*

$$\|\ln(n, m) - \sum_{p|n} \mathbf{ex}(p, n) \ln(p, m)\| = \mathcal{O}(\frac{|n|}{m})$$

Proof. First, note that the sum $\sum_{p|n} \mathbf{ex}(p, n) \ln(p, m)$ has length $\leq |n|$, and therefore is definable in $\mathbf{I\Delta}_0$. Each inequalities can be proved by induction on n using Lemma 2.5 **b**. □

2.7 The Function *lbc*

Note that

$$n! = \prod_{p \leq n} p^{e_p} \qquad \text{where} \;\; e_p = \sum_{1 \leq j \wedge p^j \leq n} \lfloor n/p^j \rfloor \tag{18}$$

We use the function **exfac** for e_p above.

Corollary 2.10. *The following function is provably total in* $\mathbf{I\Delta}_0$*:*

$$\mathbf{exfac}(p, n) = \sum_{1 \leq j \wedge p^j \leq n} \lfloor n/p^j \rfloor$$

Also, $\overline{\mathbf{I\Delta}}_0$ *proves that*

$$\mathbf{exfac}(p, 1) = 0, \qquad \text{and} \qquad \mathbf{exfac}(p, n) = \mathbf{ex}(p, n) + \mathbf{exfac}(p, n - 1) \tag{19}$$

Proof. The fact that $\mathbf{exfac}(p,n)$ is provably total in $\mathbf{I\Delta}_0$ follows from Theorem 2.3 and the fact that the sum in the definition of $\mathbf{exfac}(p,n)$ has length $\leq |n|$. The second property in (19) is proved by induction on n. $\square$

Lemma 2.11 (Provable in $\overline{\mathbf{I\Delta}_0}$)**.**

$$0 \leq \mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) \leq \frac{\ln(2n)}{\ln(p)} + \mathcal{O}(\frac{|n|}{m})$$

Proof. By definition,

$$\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) = \sum_{1\leq j \wedge p^j \leq 2n} (\lfloor \frac{2n}{p^j} \rfloor - 2\lfloor \frac{n}{p^j} \rfloor)$$

It is provable in $\mathbf{I\Delta}_0$ that

$$0 \leq \lfloor \frac{2n}{p^j} \rfloor - 2\lfloor \frac{n}{p^j} \rfloor \leq 1$$

So we have $\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) \geq 0$, and

$$\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) \leq \sum_{1\leq j \wedge p^j \leq 2n} 1 = max\{j \ : \ p^j \leq 2n\}$$

Using Lemma 2.5 we can prove by induction that

$$p^j \leq 2n \supset \|\ln(p^j) - j\ln(p)\| = \mathcal{O}(\frac{j|p|}{m})$$

It follows that

$$max\{j \ : \ p^j \leq 2n\} \leq \frac{\ln(2n)}{\ln(p)} + \mathcal{O}(\frac{|n|}{m})$$

This concludes the proof of the lemma. $\square$

Note that from (18) we have

$$\frac{(2n)!}{n!n!} = \prod_{p\leq 2n} p^{e'_p} \qquad \text{where } \ e'_p = \sum_{1\leq j \wedge p^j \leq 2n} (\lfloor 2n/p^j \rfloor - 2\lfloor n/p^j \rfloor) \tag{20}$$

Now we introduce the following functions (*lbc* stands for *logarithm of binomial coefficient*):

$$lbc(n) = \ln(\frac{(2n)!}{n!n!}) = \sum_{p\leq 2n} e'_p \ln(p) = \sum_{p\leq 2n} (\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n))\ln(p)$$

Recall that $\mathcal{P}$ denotes the set of prime numbers. The function *lbc* is formally defined as follows.

Definition 2.12. *Let lbc' be the function with the following defining axioms*

$$lbc'(n,1) = 0$$

$$lbc'(n,k+1) = \begin{cases} lbc'(n,k) & \text{if } k+1 \notin \mathcal{P} \\ lbc'(n,k) + (\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n))\ln(p) & \text{if } k+1 = p \in \mathcal{P} \end{cases}$$

Let $lbc(n) = lbc'(n,2n)$.

Theorem 2.13. *It is provable in $\overline{\mathbf{I\Delta}}_0(lbc)$ that*

$$lbc(n) = \sum_{i=1}^{2n} \ln(i) - 2\sum_{i=1}^{n} \ln(i) + \mathcal{O}(\frac{n|n|}{m})$$

Proof. We prove the theorem by induction on n. For the induction step, it suffices to show that

$$lbc(n+1) - lbc(n) = \ln(2n+1) + \ln(2n+2) - 2\ln(n+1) + \mathcal{O}(\frac{|n|}{m})$$

Using Lemma 2.5 **b**) this amounts to

$$lbc(n+1) - lbc(n) = \ln(2n+1) + \ln(2) - \ln(n+1) + \mathcal{O}(\frac{|n|}{m})$$

Thus, by Lemma 2.9 it suffices to show that

$$lbc(n+1) - lbc(n) = \sum_{p|2n+1} \mathbf{ex}(p,2n+1)\ln(p) + \ln(2) - \sum_{p|n+1} \mathbf{ex}(p,n+1)\ln(p) \tag{21}$$

By considering the cases: $p = 2$, $p|n+1$, $p|2n+1$ and $p \nmid n+1 \wedge p \nmid 2n+1$, it can be proved in $\overline{\mathbf{I\Delta}}_0$ that for all primes p,

$$\begin{aligned}&\mathbf{exfac}(p,2(n+1)) - 2\mathbf{exfac}(p,n+1) = \\ &\qquad \mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) + \mathbf{ex}(p,2n+1) + \mathbf{ex}(p,2) - \mathbf{ex}(p,n+1)\end{aligned}$$

Now (21) can be proved by proving by induction on $k \geq 2$ that

$$\begin{aligned}&lbc(n+1,k) = lbc(n,k) + \\ &\sum_{p|2n+1, p\leq k} \mathbf{ex}(p,2n+1)\ln(p) + \ln(2) - \sum_{p|n+1, p\leq k} \mathbf{ex}(p,n+1)\ln(p)\end{aligned}$$ □

2.8 Defining *lbc* in $\mathbf{I\Delta}_0(\xi)$

The theory $\mathbf{I\Delta}_0(\xi) + def(\xi)$ [WC07] is obtained from $\mathbf{I\Delta}_0$ by augmenting the function ξ and its defining axioms. The function $\xi(x) = \xi(x,y,e)$ [WC07] is

$$\xi(x) = \#\{p \ : \ p \in \mathcal{P}, p \leq x, \text{and } \lfloor y/p^e \rfloor \text{ is odd}\}$$

and has defining axioms (suppressing y, e):

$$\xi(0) = 0$$
$$\xi(x+1) = \begin{cases} \xi(x)+1 & \text{if } x+1 \in \mathcal{P} \text{ and } \lfloor y/(x+1)^e \rfloor \text{ is odd} \\ \xi(x) & \text{otherwise} \end{cases}$$

Here we show that our function *lbc* is definable in $\mathbf{I\Delta}_0(\xi) + \mathit{def}(\xi)$. As a result, the lower bounds for $\pi(n)$ and $\pi(2n)-\pi(n)$ that we prove in the following sections are also theorems of $\mathbf{I\Delta}_0(\xi)+\mathit{def}(\xi)$. Thus we obtain alternative proofs for the results from [WC07].

Theorem 2.14. *The function lbc with defining axioms given in Definition 2.12 is definable in* $\mathbf{I\Delta}_0(\xi) + \mathit{def}(\xi)$.

Proof. We show how to compute $\mathit{lbc}'(n,k)$ in $\mathbf{I\Delta}_0(\xi)$. Note that

$$\mathit{lbc}'(n,k) = \sum_{p \le k} (\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n)) \ln(p)$$

and by Lemma 2.11,

$$0 \le \mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) \le \frac{\ln(2n)}{\ln(p)} + \mathcal{O}(\frac{|n|}{m})$$

By definition,

$$\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) = \sum_{p^j \le 2n} \lfloor 2n/p^j \rfloor - 2 \sum_{p^j \le 2n} \lfloor n/p^j \rfloor$$

So, since the summations have length $\le |n|$, it is provable in $\overline{\mathbf{I\Delta}}_0$ that

$$\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) = \sum_{p^j \le 2n} (\lfloor 2n/p^j \rfloor - 2\lfloor n/p^j \rfloor)$$

In other words,

$$\mathbf{exfac}(p,2n) - 2\mathbf{exfac}(p,n) = \#\{j \le \frac{\ln(2n)}{\ln(p)} \ : \ \lfloor 2n/p^j \rfloor \text{ is odd}\}$$

As a result,

$$\mathit{lbc}'(n,k) = \sum_{j \le \ln(2n)} \left(\sum_{p \le k \wedge \lfloor 2n/p^j \rfloor \text{ is odd}} \ln(p) \right)$$

The summation in brackets can be computed in $\mathbf{I\Delta}_0(\xi)$ using the counting function ξ just as described in Theorem 2.6 and Theorem 2.7. $\square$

3 A Lower Bound for $\pi(n)$ in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$

Note that $\pi(2n-1) = \pi(2n)$ for $n \geq 2$. So it suffices to give a lower bound for $\pi(2n)$. We choose a simple proof for the $\Omega(n/\ln(n))$ lower bound for $\pi(2n)$ and point out that this proof can be formalized using the function *lbc* introduced above. From this lower bound for $\pi(n)$ we can derive in $\mathbf{I\Delta}_0(\pi, lbc')$ the fact that there are infinitely many prime numbers.

The idea is to compute an upper bound and a lower bound for $\frac{(2n)!}{n!n!}$; by comparing these bounds we can derive a lower bound for $\pi(2n)$. In our formalization, we will use $lbc(n)$ instead of $\frac{(2n)!}{n!n!}$.

Lemma 3.1 (Provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$)**.**

$$lbc(n) \leq \pi(2n)(\ln(2n) + \mathcal{O}(\frac{|n|}{m}))$$

Proof. We prove by induction on $k \leq 2n$ that $lbc'(n,k) \leq \pi(k)\ln(2n)$ using the defining axioms for lbc' (Definition 2.12) and Lemma 2.11. □

Lemma 3.2 (Provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$)**.** *For $n > m$:*

$$lbc(n) = 2n\ln(2) + c(m) + \mathcal{O}(\frac{n|n|}{m}) \tag{22}$$

for some constant $c(m)$ depends only on m.

Proof. By (10) in Theorem 2.6 we have

$$\sum_{i=1}^{2n} \ln(i) - \sum_{i=1}^{n} \ln(i) = (2n + (h-2)2^{h+1})\ln(2) + T - S$$

where T, S depend only on m (recall also that $m = 2^h$). Now the lemma follows from Theorem 2.13. □

Corollary 3.3 (Provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$)**.**

$$\pi(n) = \Omega(n/\ln(n)) \tag{23}$$

It follows that the existence of arbitrarily large prime numbers is provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$.

4 Bertrand's Postulate and a Lower Bound for $\pi(2n) - \pi(n)$

We will prove Bertrand's Postulate (that $\pi(2n) - \pi(n) \geq 1$ for all n) and a lower bound for the number of prime numbers between n and $2n$: $\pi(2n) - \pi(n) = \Omega(n/\ln(n))$. For the latter, we follow the proof from [Mos49]. First we outline the proof of the lower bound for $\pi(2n) - \pi(n)$; the formalizations are given in Section 4.1.

Recall Chebyshev's function $\vartheta(x)$ from (14).

Theorem 4.1. *For $n \geq 1$, $\vartheta(n) < 2n\ln(2)$.*

Proof. First, because
$$\frac{(2k+1)!}{k!(k+1)!}$$
appears twice in the binomial expansion of 2^{2k+1}, we have
$$\frac{(2k+1)!}{k!(k+1)!} \leq \frac{1}{2}2^{2k+1} = 2^{2k} \tag{24}$$
Also, all primes p where $k+1 < p \leq 2k+1$ divide $\frac{(2k+1)!}{k!(k+1)!}$. Hence
$$\prod_{k+1<p\leq 2k+1} p \leq \frac{(2k+1)!}{k!(k+1)!} \tag{25}$$
Consequently,
$$\vartheta(2k+1) - \vartheta(k+1) = \sum_{k+1<p\leq 2k+1} \ln(p) \leq \ln\frac{(2k+1)!}{k!(k+1)!} \leq \ln(2^{2k}) = 2k\ln(2) \tag{26}$$
Now we prove the theorem by induction on n. The base cases ($n = 1$ and $n = 2$) are trivial. For the induction step, the case where n is even is also obvious, since then $\vartheta(n) = \vartheta(n-1)$. So suppose that $n = 2k+1$. Using (26) and the induction hypothesis (for $n = k+1$) we have $\vartheta(2k+1) < 2k\ln(2)+2(k+1)\ln(2) = 2(2k+1)\ln(2)$. □

Note that this theorem gives a $\mathcal{O}(n/\ln(n))$ upper bound for $\pi(n)$, but we do not need this fact here.

Lemma 4.2.
$$\frac{(2n)!}{n!n!} \leq (2n)^{\sqrt{2n}} \left(\prod_{\sqrt{2n}<p\leq 2n/3} p\right)\left(\prod_{n<p<2n} p\right) \tag{27}$$

Proof. From (20), by noting that
$$e'_p \begin{cases} = 1 & \text{if } n < p < 2n \\ = 0 & \text{if } 2n/3 < p \leq n \\ \leq 1 & \text{if } \lceil\sqrt{2n}\rceil \leq p \leq \lfloor 2n/3 \rfloor \\ \leq \frac{\ln(2n)}{\ln(p)} & \text{if } p < \sqrt{2n} \end{cases}$$
□

Corollary 4.3. $\pi(2n) - \pi(n) = \Omega(n/\ln(n))$.

Proof. Note that
$$\frac{(2n)!}{n!n!} \geq \frac{2^{2n}}{2n+1}$$

(because $\frac{(2n)!}{n!n!}$ is the largest coefficient in $(1+1)^{2n}$). Therefore

$$\ln(\frac{(2n)!}{n!n!}) \geq 2n\ln(2) - \ln(2n+1)$$

Also,

$$\ln\left(\prod_{\sqrt{2n}<p\leq 2n/3} p\right) \leq \ln\left(\prod_{p\leq 2n/3} p\right) = \vartheta(2n/3)$$

so by Theorem 4.1,

$$\ln\left(\prod_{\sqrt{2n}<p\leq 2n/3} p\right) < 4n\ln(2)/3$$

In addition,

$$\ln\left(\prod_{n<p<2n} p\right) < (\pi(2n) - \pi(n))\ln(2n)$$

As a result, by taking logarithm of both sides of (27) we have

$$2n\ln(2) - \ln(2n+1) < \sqrt{2n}\ln(2n) + 4n\ln(2)/3 + (\pi(2n) - \pi(n))\ln(2n)$$

From this the conclusion follows easily. □

4.1 Formalization in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$

Recall (Section 2.4) that our version of Chebyshev's function, $\vartheta(x,m)$, or simply $\vartheta(x)$, is definable in $\mathbf{I\Delta}_0(\pi)$. Following Theorem 4.1 we prove:

Theorem 4.4 (Provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$)**.** *For some constant $c'(m)$,*

$$\vartheta(n,m) \leq 2n\ln(2) + |n|c'(m) + \mathcal{O}(\frac{n|n|}{m})$$

Proof. Note that

$$\ln(\frac{(2k+1)!}{k!(k+1)!}) = lbc(k+1) - \ln(2)$$

Using Lemma 2.9 and from the definition of lbc (Definition 2.12), we can prove in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$ that

$$\ln(2) + \sum_{k+1<p\leq 2k+1} \ln(p) \leq lbc(k+1)$$

(By proving by induction on $j \leq 2k$ that

$$\ln(2) + \sum_{k+1<p\leq j} \ln(p) \leq lbc'(k+1, j)$$

We will have to consider two cases: either $k+1$ is a power of 2, or not.)

As a result, by Lemma 3.2 we have

$$\sum_{k+1<p\leq 2k+1} \ln(p) \leq lbc(k+1) - \ln(2) = 2k\ln(2) + (c(m)+\ln(2)) + \mathcal{O}(\frac{k|k|}{m})$$

That is, for $c'(m) = c(m) + \ln(2)$,

$$\vartheta(2k+1) - \vartheta(k+1) \leq 2k\ln(2) + c'(m) + \mathcal{O}(\frac{k|k|}{m})$$

Now we can prove by strong induction on k that

$$\vartheta(k) \leq 2k\ln(2) + |k|c'(m) + \mathcal{O}(\frac{k|k|}{m})$$

(using the fact that $|2k+1| = |k|+1$). □

Following Lemma 4.2 we have:

Lemma 4.5 (Provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$)**.**

$$lbc(n) \leq \lfloor\sqrt{2n}\rfloor \ln(2n) + \vartheta(\frac{2n}{3}) + \sum_{n<p<2n} \ln(p)$$

Proof. The proof is similar to the proof of Lemma 3.1. First we prove by induction on k that

$$lbc'(n,k) \leq \pi(k)\ln(2n)$$

for $k \leq \lfloor\sqrt{2n}\rfloor$. Then we prove by induction on k, where $\lfloor\sqrt{2n}\rfloor < k \leq \frac{2n}{3}$, that

$$lbc'(n,k) \leq \lfloor\sqrt{2n}\rfloor \ln(2n) + \vartheta(k)$$

Finally, we prove by induction on k where $n < k < 2n$ that

$$lbc'(n,k) \leq \lfloor\sqrt{2n}\rfloor \ln(2n) + \vartheta(\frac{2n}{3}) + \sum_{n<p\leq k} \ln(p)$$ □

Corollary 4.6 (Provable in $\overline{\mathbf{I\Delta}}_0(\pi, lbc')$)**.**

$$\pi(2n) - \pi(n) = \Omega(\frac{n}{\ln(n)})$$

Proof. By Lemma 3.2, Theorem 4.4 and the above lemma we have

$$2n\ln(2) + c(m) + \mathcal{O}(\frac{n|n|}{m}) \leq \lfloor\sqrt{2n}\rfloor \ln(2n) + \left(\frac{4n\ln(2)}{3} + |\frac{2n}{3}|c'(m) + \mathcal{O}(\frac{n|n|}{m})\right) + \sum_{n<p<2n} \ln(p)$$

It follows that for $n > m^2, m > |n|^2$:

$$\sum_{n<p<2n} \ln(p) \geq \frac{2\ln(2)}{3} n - \mathcal{O}(\frac{n|n|}{m})$$

The conclusion follows from the fact provable in $\overline{\mathbf{I\Delta}}_0(\pi)$ that the LHS is at most $(\pi(2n) - \pi(n))\ln(2n)$. □

Corollary 4.7 (Provable in $\mathbf{I\Delta}_0(\pi, \mathit{lbc}')$)**.** *For all n, $\pi(2n) - \pi(n) \geq 1$.*

Proof. The previous corollary shows that for some standard threshold $n_0 \in \mathbb{N}$, $\pi(2n) - \pi(n) > 0$ for all $n \geq n_0$. The fact that $\pi(2n) - \pi(n) \geq 1$ for $n < n_0$ is true in $\mathbb{N}$, and hence is provable in $\mathbf{I\Delta}_0$. □

5 Conclusion

Sylvester's Theorem asserts that for $1 \leq x \leq y$, some number among

$$y + 1, y + 2, \ldots, y + x$$

has a prime divisor $p > x$. In [Woo81] it is shown that Sylvester's Theorem can be proved in $\mathbf{I\Delta}_0 + \mathit{PHP}(\mathbf{\Delta}_0)$. Here, as well as in [Cor95, WC07], we have a $\Omega(n/\ln(n))$ lower bound for $\pi(2n)-\pi(n)$, the number of prime numbers between n and $2n$. Such lower bound does not seem to follow from the proof in [Woo81]. However, it is not clear whether $\mathit{PHP}(\mathbf{\Delta}_0)$ is provable in $\mathbf{I\Delta}_0(\pi, \mathit{lbc}')$ or even $\mathbf{I\Delta}_0(\xi) + \mathit{def}(\xi)$.

Also, as far as we know, the axiom for *lbc* considered here (or even the axiom for ξ considered in [WC07]) and the axiom for K [Cor95] are incomparable over $\mathbf{I\Delta}_0(\pi)$. It is an interesting problem to see whether one follows from the other in $\mathbf{I\Delta}_0$.

Acknowledgments: I would like to thank Steve Cook and the referees for their helpful comments.

References

[Ben62] James Bennett. *On Spectra.* PhD thesis, Princeton University, Departmentof Mathematics, 1962.

[Bus98] Samuel Buss. First–Order Proof Theory of Arithmetic. In S. Buss, editor, *Handbook of Proof Theory*, pages 79–147. Elsevier, 1998.

[CD94] C. Cornaros and C. Dimitracopoulos. The Prime Number Theorem and Fragments of **PA**. *Archive for Mathematical Logic*, 33:265–281, 1994.

[CN06] Stephen Cook and Phuong Nguyen. Foundations of Proof Complexity: Bounded Arithmetic and Propositional Translations. Book in progress, 2006.

[Coo07] Stephen Cook. Bounded Reverse Mathematics. Plenary Lecture for CiE 2007, 2007.

[Cor95] Ch. Cornaros. On Grzegorczyk Induction. *Annals of Pure and Applied Logic*, 74:1–21, 1995.

[HP93] Petr Hájek and Pavel Pudlák. *Metamathematics of First-Order Arithmetic*. Springer–Verlag, 1993.

[Kra95] Jan Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory*. Cambridge University Press, 1995.

[Mos49] Leo Moser. A theorem on the distribution of primes. *American Mathematical Monthly*, 56(9):624–625, 1949.

[NC05] Phuong Nguyen and Stephen Cook. Theory for $\mathbf{TC}^0$ and Other Small Complexity Classes. *Logical Methods in Computer Science*, 2, 2005.

[Ngu08] Phuong Nguyen. *Bounded Reverse Mathematics*. PhD thesis, University of Toronto, 2008. `http://www.cs.toronto.edu/~pnguyen/`.

[PWW88] J.B. Paris, A.J. Wilkie, and A.R. Woods. Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, 53(4):1235–1244, 1988.

[WC07] Alan Woods and Ch. Cornaros. On bounded arithmetic augmented by the ability to count certain sets of primes. Unpublished, 2007.

[Woo81] Alan Woods. *Some Problems in Logic and Number Theory and Their Connections*. PhD thesis, University of Manchester, 1981.