The calculation of $p_n$ and $\pi(n)$

Simon Plouffe
June 03, 2020

Abstract

A new approach is presented for the calculation of $p_n$ and $\pi(n)$ which uses the Lambert W function. An approximation is first found and using a calculation technique it makes it possible to have an estimate of these two quantities more precise than those known from Cipolla and Riemann. The calculation of $p_n$ uses an approximation using the Lambert W function and an estimate based on a logarithmic least square curve (LLS) $c(n)$. The formula is:

$$p_n + \pi(n) \approx -nW_{-1} \left( \frac{-e}{n} \right) c(n)$$

The results presented are empirical and apply up to $n \approx 1.358 \times 10^{16}$. 
Introduction

Today we know very large prime numbers like $2^{2^{82589933}} - 1$ but the rank of these numbers is unknown. The primes with their rank is known up to $10^{27}$ for $\pi(n)$ and $10^{24}$ for $p_n$ at some specific points and the complete list of primes with their rank is known only up to $10^{17}$.

In 2010, Dusart [16] proved that $\pi(n) \approx \frac{n}{\ln(n) - 1}$ if $n > 5393$. We will use this approximation to give one approximation of $p_n$ by inverting the formula.

If $\pi(n) \approx \frac{n}{\ln(n) - 1}$ then $p_n \approx -nW_{-1}\left(\frac{-e}{n}\right)$.

$W_{-1}(n)$ is the Lambert W function of order -1, $W_0$ or $W(n)$ is of order 0. We take the branch of the W function which gives meaning to quantity $\frac{-e}{n}$.

That formula is quite precise, for $n = 10^{24}$, we have $p_n$ precise to 99.97 %.

By analyzing the rest of $p_n$ and $-nW_{-1}\left(\frac{-e}{n}\right)$, we rapidly find that it is close to $\frac{n}{W(n)}$, so

$p_n \approx -nW_{-1}\left(\frac{-e}{n}\right) - \frac{n}{W(n)}$. But also that the quantities

$$\frac{n}{\ln(n) - 1} \approx Li(n) \approx \pi(n) \approx \frac{n}{W(n)} \approx \frac{\rho_n}{2\pi}$$

are close too when $n \to \infty$. Here, $\rho_n$ is the imaginary part of the n'th non-trivial zero of Riemann's $\zeta$ and $Li(n)$ is the integral logarithm, the values are: .1843e23, .18434e23, .1844e23, .1948e23 et .1986e23 respectively.

Empirical data suggest that the choice of $\pi(n)$ is the best.

If $n = 10^{24}$ then

$$p_{10^{24}} + \pi(10^{24}) \approx -10^{24}W_{-1}\left(\frac{-e}{10^{24}}\right)$$

(2)

The approximation is valid up to 0.99999401 or 99.999401 % of the real value.

A naive question is then, couldn’t we be more precise? Generally speaking, is there a way to have an exact formula? If we can be more precise then how much? What about the rest? What is his nature exactly?

An explicit formula: Prime numbers in geometric progression

Right now, the question of whether an exact formula exists for prime numbers is yes and no both. There are many formulas that give them but they are either impractical or limited see the table in appendix.

For example, here is an explicit formula which gives some primes in geometric progression.

$$f(n) = \{c^n\}$$

(3)
Here \{\} is the nearest integer function. The smallest example of \(c\) is: if \(c = 2.553854696...\) then \(f(n)\) is 3, 7, 17, 43, 109, 277, 709. The sequence contains only 7 terms. But can we go further? Using the simulated annealing method and Monte Carlo we can go much further.

<table>
<thead>
<tr>
<th>Constant (c) such that ({c^n}) is prime</th>
<th>Values</th>
<th>Number of primes generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.553854696...</td>
<td>3, 7, 17, 43, 109, 277, 709</td>
<td>7</td>
</tr>
<tr>
<td>593.46526943871...</td>
<td>(n=2.48)</td>
<td>47</td>
</tr>
<tr>
<td>2027.1671684764912194343956</td>
<td>(n=1.97)</td>
<td>97</td>
</tr>
<tr>
<td>29983.279826631136...</td>
<td>(n=1.422)</td>
<td>422</td>
</tr>
<tr>
<td>55237.07504296764715433124781528617...</td>
<td>(n=2.633)</td>
<td>632</td>
</tr>
</tbody>
</table>

From there we can conjecture that if \(c\) is large enough, the sequence with formula (3) can generate an arbitrary number of primes explicitly, see [23].

All we have for the expression of \(p_n\) or \(\pi(n)\) for large values of \(n\) are just approximations. The best known approximation is that of Riemann, the function is called Riemann-R or that of Gram found in 1884. The only way that has been found is to use partially Gram's formula followed by a sophisticated Erathosthenes screen. This is why that the value of \(p_n\) or \(\pi(n)\) is only known up to \(10^{24}\) and \(10^{27}\) respectively. The last calculation of \(\pi(10^{27})\) took several months and a significant amount of computer time (23 CPU years) in 2018.

**First approximation**

We will confine ourselves for the moment to the calculation of \(p_n\) since \(\pi(n)\) can be calculated by inversion.

The classic formula for \(p_n\) is \(p_n \approx n \ln(n)\) or better yet the one found by Cipolla in 1902 states that

\[
p_n \approx n \left( \ln(n) + \ln(\ln(n)) - 1 + \frac{\ln(\ln(n)) - 2}{\ln(n)} - \frac{(\ln(\ln(n)))^2 - 6 \ln(\ln(n)) + 1}{2 \ln(n)^2} + \cdots \right).
\]  

The calculation was taken further in 1994 by B. Salvy who extracted a procedure for pushing it further the approximation.

What is remarkable is the similarity with the asymptotic development of \(W(n)\).

\[
W(n) \approx L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(-2 + L_2)}{2L_1^2} + \frac{L_2(6 - 9L_2 + 2L_2^2)}{6L_1^3} + \frac{L_2(-12 + 36L_2 - 22L_2^2 + 36L_2^3)}{12L_1^4} + \cdots
\]

\(L_1 = \ln(n)\) et \(L_2 = \ln(\ln(n))\).

The calculation of \(p_n\) with this formula (4) gives 12 exact decimal places out of the 26 that count \(p_{10^{24}}\). But does not allow to go further, even with 64 terms in the asymptotic development. The formula of \(Gram^{<-1>}\) is much more precise. \(Gram^{<-1>}\) is the functional inverse of the Gram series [24].
A better approximation

A summary analysis indicates that the rest after the first term of (1) is of type logarithmic: \( a + b \ln(n) \) where \( a \) and \( b \) to be determined. An idea is then to calculate the logarithmic least squares curve (LLS) passing through a chosen number of points on a table of values of primes.

We can also notice that by taking only one term for the approximation of \(-nW_{-1}\left(\frac{-e}{n}\right)\), this form is equivalent to several terms of Cipolla’s development. If we take the 2 terms it will be even more precise. In other words, given the nature of the asymptotic development of \( W(n) \), each term is equivalent to several terms in the development of Cipolla. We assume here that the remainder after the 2 terms is a logarithmic curve and that once calculated it will stick to reality.

The question then arises of what is the nature of what remains? In fact, we do not know. The best known for \( \pi(n) \) which is theoretically valid is \( Li(n) \). Riemann proposed a 2nd formula that is much better at first but was invalidated by Littlewood in 1914. This 2nd formula, called Riemann R or equivalent, the Gram series is

\[
\pi(n) = \sum_{k=1}^{\infty} \frac{\ln(x)^k}{k! \zeta(k + 1)}
\]

Numerically, the approximation of \( \pi(n) \) by Riemann’s R or the Gram series is excellent in addition to converging quickly. But Littlewood showed that after \( 10^9 \), the approximation is less precise. As for the function \( Li(n) \), it behaves better at much larger scales, the first crossover of \( Li(n) \) and \( \pi(n) \) was evaluated at \( 10^{316} \) approximately, that is \( Li(n) - \pi(n) = 0 \) in the neiboughoud of \( 1.397 \times 10^{316} \).

The best approximation that has been found empirically is:

\[
p_n = -nW_{-1}\left(\frac{-e}{n}\right) c(n) - \pi(n)
\]

where \( c(n) \) is of type \( a + b \log(n) \).

Here is a table of values that were considered for the calculation of \( p_n \) from \( 10^2 \) to \( 10^{17} \):

<table>
<thead>
<tr>
<th>Value of the step</th>
<th>Number of values</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>27117419</td>
<td>2711741900</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>32082085</td>
<td>32082085000</td>
</tr>
<tr>
<td>( 10^4 )</td>
<td>45020269</td>
<td>450202690000</td>
</tr>
<tr>
<td>( 10^5 )</td>
<td>16038989</td>
<td>1.603 \times 10^{12}</td>
</tr>
<tr>
<td>( 10^6 )</td>
<td>4046531</td>
<td>4.046531 \times 10^{12}</td>
</tr>
<tr>
<td>( 10^7 )</td>
<td>5011691</td>
<td>5.011691 \times 10^{13}</td>
</tr>
<tr>
<td>( 10^8 )</td>
<td>454060</td>
<td>4.54060 \times 10^{13}</td>
</tr>
<tr>
<td>( 10^9 )</td>
<td>2200000</td>
<td>2.2 \times 10^{15}</td>
</tr>
<tr>
<td>( 10^{10} )</td>
<td>1358121</td>
<td>1.358121 \times 10^{16}</td>
</tr>
<tr>
<td>( 10^{11} )</td>
<td>135812</td>
<td>1.358121 \times 10^{16}</td>
</tr>
<tr>
<td>( 10^{12} )</td>
<td>54974</td>
<td>5.4974 \times 10^{16}</td>
</tr>
</tbody>
</table>
We can calculate the points on the LLS curve by solving the following equation for $x$, taking $n$ in one of the tables above.

$$-nW_{-1}\left(\frac{-e}{n}\right)x - \pi(n) + p_n = 0$$

by using the bisection method. The values are between 0.9 and 1. We then calculate the logarithmic least squares curve. The coefficient $r^2$ will indicate if the curve is right. The calculation of coefficients $a$ and $b$ is done according to the formula:

$$b = \frac{n \sum_{i=1}^{n} (y_i \ln x_i) - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} \ln x_i}{n \sum_{i=1}^{n} (\ln x_i)^2 - (\sum_{i=1}^{n} \ln x_i)^2}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} (\ln x_i)}{n}$$

Recall, the coefficient $r^2$ indicates whether the experimental data stick to the curve. If is near 1 or 0, the curve follows a straight line very closely. The logarithmic least-squares curve is simply the log of values that are aligned on a line.

The curve thus found for $n \leq 1.358121 \times 10^{16}$ is

$$c(n) = 1.0000314775792421150615325693061 - 0.00000051483940138413674623044640769440 \ln(n)$$

Once you have found this LLS curve, you can significantly increase the accuracy if you use a trick. We take the chosen interval, $n \cdot 10^{10}$, $n = 1..1358121$. We then separate into 1358 slices of 10,000 values and for each value we modify the curve $c(n)$ with the formula

$$c(n) \rightarrow c(n) s^k$$

And $s = (1 - 10^{-10})$, then it suffices to find for which $k$, the curve admits a minimum error if we compare to the true values of $p_n$. An experiment was conducted with 13581 intervals to see if the mean value of the deviation decreased: this is the case. The reasonable limit found here is from 1,358 slices of 10,000 values.

In the appendix you can consult the program written in Maple language which performs the operation.

### Calculation of $\pi(n)$

To calculate $\pi(n)$, it suffices to isolate $\pi(n)$ in the equation of $G(n)$, normally trivial but in practical it does not work. Indeed $\pi(n)$ is smaller than $p_n$ in size. The formula remains valid except

<table>
<thead>
<tr>
<th>$n$</th>
<th>$c(n)$</th>
<th>$d(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{13}$</td>
<td>12317</td>
<td>$1.2317 \times 10^{17}$</td>
</tr>
<tr>
<td>$10^{14}$</td>
<td>2162</td>
<td>$2.162 \times 10^{17}$</td>
</tr>
</tbody>
</table>
that the coefficients change slightly. The calculation of can be done with formula 1, with the primecount program which gives it directly. We can also have the values of  \( \text{Gram}(n) \) and \( \text{Gram} \) inverted. If instead of \( \pi(n) \) in formula (1) we rather take the term \( \frac{n}{W(n)} \), it’s a little less precise but still allows good accuracy.

For example, for the interval from 1 to 1000000 the following program calculates \( p_n \) very precisely.

\[
G:=\text{proc}(n) \text{# calculation of } p(n) \text{ up to 1 million (first million of primes).}
\text{local ll, lk, s, s2, ss, k;}
\]
\[
\text{ll} := [82, 16, -14.4, -31, -10, -31, -32, -1, -44, -17, -38, -8, 7, -35, -41, -36, -14, -27, -12, -5, -51, -40, 17, 7, -16, -14, 13, -5, -1, -26, -9, -27, -31, -9, 8, 16, 4, 9, 0, 20, 11, 7, -15, -23, -17, -10, -2, -5, 8, 7, 9, 3, -12, -11, 4, 5, -3, 9, -1, 7, 24, 25, 33, 20, 15, 11, 9, 3, 9, 15, 3, 1, 4, 5, -9, -1, -12, -4, 14, 16, 17, 28, 18, 12, 21, 24, 10, 14, 16, 15, 24, 26, 36, 30];
\]
\[
lk := \text{floor}(1/10000*n) + 1;
\]
\[
s := 0.3238679016803340 + 0.4042167153029803e-1*\text{ln}(kk):
\]
\[
s2 := \text{subs}(kk = n, s):
\]
\[
ss := s2*0.999^ll[0]
\]
\[
\text{round(abs(-n*W(-1, -exp(1.0)/n)) - ss*fab(n/W(n))))};
\]
\end;
\]

By inverting \( \pi(n) \) to find \( \pi(n) \) over the interval 1 to 135000x10e16. We have the following program. It has the advantage of not having to calculate or having a table of the values of the first, but it is less precise.

\[
f:=\text{proc}(k) \text{ local n,a,bas,haut, ll,lk,s,ss; # calcul de } \pi(n) \text{ jusqu’a } 10^16
\]
\[
haut:=\text{evalf}(2*k/\log(k));
\]
\[
bas:=\text{haut/8};
\]
\[
ll:=[3122, 3186, 3222, 3243, 3251, 3270, 3272, 3282, 3289, 3291, 3289, 3297, 3312,
3305, 3300, 3313, 3308, 3311, 3321, 3318, 3323, 3322, 3326, 3322, 3319, 3322,
3328, 3328, 3327, 3335, 3334, 3336, 3336, 3337, 3335, 3335, 3331, 3334, 3337,
3334, 3341, 3343, 3345, 3346, 3343, 3340, 3341, 3342, 3348, 3348, 3347, 3349,
3349, 3345, 3349, 3350, 3348, 3352, 3351, 3346, 3344, 3344, 3348, 3348,
3349, 3354, 3355, 3357, 3355, 3354, 3355, 3358, 3359, 3358, 3358, 3358, 3357,
3357, 3358, 3356, 3355, 3357, 3358, 3359, 3358, 3358, 3355, 3355, 3360, 3357, 3354,
3358, 3358, 3362, 3361, 3360, 3360, 3361, 3360, 3362, 3361, 3362, 3363,
3363, 3362, 3364, 3363, 3363, 3361, 3360, 3361, 3364, 3366, 3366, 3366,
3366, 3369, 3369, 3368, 3366, 3367, 3366, 3368, 3368, 3368, 3368, 3366,
3366, 3365, 3364, 3365, 3365];
\]
\[
lk:=\text{floor}(k/(10000000000000000)+1):
\]
\[
s:=0.882819461483173314372633 + 0.855943969794036445417381e-3*\text{ln}(n):
\]
\[
ss:=s*(0.999999)*ll[ll[0]]:
\]
\[
a:=\text{abs(}\text{evalf}(-n*W(-1, -\text{exp}(1.0)/n))) - \text{evalf}(ss)*\text{fab}(n/W(n))-k;
\]
\[
\text{fsolve}(a,n,n=\text{haut, fulldigits});
\]
\end;
Appendix (Tables and Programs)

Calculation of $p_n$

Comparison with the range $10000000000 \times 10^{10} \sim 1.352 \times 10^{16}$

<table>
<thead>
<tr>
<th>Formula for $p_n$</th>
<th>Formula of Gram inverted</th>
<th>$G(n)$</th>
<th>Formula of Cipolla-Salvy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal gap</td>
<td>57</td>
<td>13</td>
<td>640495</td>
</tr>
<tr>
<td>Maximal gap</td>
<td>117539110</td>
<td>412614395</td>
<td>1103 millions</td>
</tr>
<tr>
<td>Average gap</td>
<td>79.23 millions</td>
<td>18.81 millions</td>
<td>510 millions</td>
</tr>
</tbody>
</table>

Conclusion: The formula with LambertW function is 4.21 times more precise than that of Gram series (inverted).

lk := fkoń/(1/10000000000000*n) + 1;
s := 1.0000314775792421150615325693061 + 0.514839401384136746230447669440*(1/1000000)*ln(kk);
s2 := subs(kk = n, s);
ss := s^2*0.9999999999^ll(10*s);
abs(-n*W(-1, -exp(1.0)/n)) - piofn
end:

Tables of primes:
http://plouffe.fr/NEW/list_primes_pi_of_n_100000000000.txt
http://plouffe.fr/NEW/list_primes_10000000000.txt

Example:
g(1327460000000000,39285023244530) = 49668015014179465.522289485977202
The real value of \( p_{1327460000000000} \) = 49668015014179453

Maple program for the calculation of \( p_n \), \( n = 1.356 \times 10^{16} \ldots 10^{24} \)

F := proc(n, piofn)
local cn, z, pola, polb;
pola := 0.18031788297753836802559072260225588343254e-12*x^8 - 0.32068536427839673078154416271278702381e-10*x^7 + 0.2521168696363102117361245200766645862014e-8*x^6 - 0.114824566104216214093938301036666192296e-6*x^5 + 0.3329760033274279782321163791428487967963e-5*x^4 - 0.643395542494949120689514623217102223176e-4*x^3 + 0.78518018575336382778142511770195187519581e-3*x^2 - 0.5912431390595954878523745751806703967872e-2*x + 0.2187329700777127284427407021768653056673e-1;
polb := 0.8949926057969637729777538473173261408730e-12*x^8 - 0.161179595080641630416491053953385128968e-9*x^7 + 0.128754231981049792998526211011490785070e-7*x^6 - 0.5988056104228871471776180438025688273194e-6*x^5 + 0.17869159102510734370249783252030617773e-4*x^4 - 0.354856585455696465095977029495597659212e-3*x^3 + 0.469042702380857960299344427037051784331e-2*x^2 - 0.3980636249963806490254767914782205838276e-1*x + 0.199974737812788247127674632264178712585;
z := evalf(log10(n));
cn := subs(x = z, pola) + subs(x = z, polb)*ln(n);
evalf(-cn*n*W(-1, -exp(1.0)/n)) - piofn
end:

<table>
<thead>
<tr>
<th>n</th>
<th>F(n)</th>
<th>p_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>10^{16}</td>
<td>39490691390373528.99999995710593</td>
<td>39490691390373529</td>
</tr>
<tr>
<td>10^{17}</td>
<td>4185296581467695668.9998280338750</td>
<td>4185296581467695669</td>
</tr>
<tr>
<td>10^{18}</td>
<td>44211790234832169331.000076399063</td>
<td>44211790234832169331</td>
</tr>
<tr>
<td>10^{19}</td>
<td>46567546511660765549.000000497931</td>
<td>46567546511660765549</td>
</tr>
<tr>
<td>10^{20}</td>
<td>4892055594575155744537.0000098572</td>
<td>4892055594575155744537</td>
</tr>
<tr>
<td>$10^{21}$</td>
<td>$51271091498016403471852.999978699$</td>
<td>$51271091498016403471853$</td>
</tr>
<tr>
<td>$10^{22}$</td>
<td>$536193870744162118627429.00001989$</td>
<td>$536193870744162118627429$</td>
</tr>
<tr>
<td>$10^{23}$</td>
<td>$5596564467986980643073682.9999696$</td>
<td>$5596564467986980643073683$</td>
</tr>
<tr>
<td>$10^{24}$</td>
<td>$58310039994836584070534263.000118$</td>
<td>$58310039994836584070534263$</td>
</tr>
<tr>
<td>n \times 10^5</td>
<td>\pi(n)</td>
<td>p_n</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>5</td>
<td>15.707963269</td>
<td>15.707963269</td>
</tr>
<tr>
<td>6</td>
<td>18.849590168</td>
<td>18.849590168</td>
</tr>
</tbody>
</table>

Table of values of n, \pi(n), p_n, G(n)
## Table of known programs, formulas for the calculation of primes.

<table>
<thead>
<tr>
<th>Author (s)</th>
<th>Year</th>
<th>Comment</th>
<th>Efficiency</th>
<th>Number of terms calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erathostène</td>
<td>-276 to -194</td>
<td>Sieve</td>
<td>Practical</td>
<td>Calculable infinity</td>
</tr>
<tr>
<td>Mersenne</td>
<td>1536</td>
<td>Primes of the form: $2^p-1$.</td>
<td>Practical and exact</td>
<td>51</td>
</tr>
<tr>
<td>Fermat</td>
<td>1640</td>
<td>Little theorem of Fermat</td>
<td>Produces weak probable primes</td>
<td>Calculable infinity</td>
</tr>
<tr>
<td>Euler</td>
<td>1772</td>
<td>Second degree polynomial</td>
<td>Practical</td>
<td>40</td>
</tr>
<tr>
<td>Mills</td>
<td>1947</td>
<td>Double exponential</td>
<td>Practical</td>
<td>Less than 10 terms</td>
</tr>
<tr>
<td>Wright</td>
<td>1951</td>
<td>Super exponential</td>
<td>Practical</td>
<td>Less than 5 terms</td>
</tr>
<tr>
<td>Wilson</td>
<td>Circa 1780</td>
<td>Formula with $p!$</td>
<td>Theoretical</td>
<td>Very few</td>
</tr>
<tr>
<td>Jones, Sato, Wada, Wiens</td>
<td>1976</td>
<td>Polynôme de degré 25 à 26 variables</td>
<td>Theoretical</td>
<td>Very few</td>
</tr>
<tr>
<td>John H. Conway</td>
<td>1987</td>
<td>FRACTRAN</td>
<td>Theoretical</td>
<td>Very few</td>
</tr>
<tr>
<td>Rowland</td>
<td>2008</td>
<td>Récurrence</td>
<td>Theoretical</td>
<td>Very few</td>
</tr>
<tr>
<td>Dress, Landreau</td>
<td>2010</td>
<td>6th degree polynomial</td>
<td>Practical</td>
<td>58</td>
</tr>
<tr>
<td>Benoit Perichon (et al.)</td>
<td>2010</td>
<td>26 primes in arithmetic progression</td>
<td>Practical</td>
<td>26</td>
</tr>
<tr>
<td><strong>Kim Walisch</strong></td>
<td>2020</td>
<td>Primesieve and Primecount</td>
<td>Practical</td>
<td>Calculable infinity</td>
</tr>
</tbody>
</table>
Bibliography (not sorted).

http://plouffe.fr/NEW/list_primes_10000000000000.txt