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

Simon Plouffe
Feb. 23 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 function $c(n)$ is the same in both cases. The two formulas are:

\[
p_n \approx -nW_1\left(\frac{-e}{n}\right) - \frac{n c(n)}{W_0(n)}
\]

\[
\pi(n) \approx \left\{-nW_1\left(\frac{-e}{n}\right) - \frac{n c(n)}{W_0(n)}\right\}^{-1}
\]

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

Introduction

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

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

This formula is precise, for $n = 10^{24}$, we have $p_n$ precise at 99.9 %.

By analyzing the remainder of $p_n$ and $-nW_{-1}\left(\frac{-e}{n}\right)$, we quickly find that it is close to $\frac{n}{W(n)}$, therefore $p_n \approx -nW_{-1}\left(\frac{-e}{n}\right) - \frac{n}{W(n)}$.

Here $W(n)$ is the Lambert W function of order 0.

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

\[
p_n \sim n \left(\ln(n) + \ln(\ln(n)) - 1 + \frac{\ln(\ln(n)) - 2}{\ln(n)} - \frac{\ln(\ln(n))^2 - 6 \ln(\ln(n)) + 11}{2 \ln(n)^2} + \ldots\right).
\]
The calculation was taken further in 1994 with Salvy who extracted a procedure from which the approximation could be taken further.

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

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

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

Again, if we analyze the rest with respect to the true value of \( p_n \) we find with \( n = 10^{24} \) (this is the best known value of \( p_n \)).

\[
p_n \approx 58308642550474983476717666
\]

The real value being 58310039994836584070534263, we therefore have an approximation to 0.999976, that is to say 99.9976%. We therefore gained 2 orders of magnitude.

**A better approximation**

A summary analysis indicates that the remainder after the 2 terms \( W_{-1}(\frac{e}{n}) \) and \( W(n) \) is logarithmic in nature. A simple idea is then to calculate the logarithmic least squares curve or LLS curve. We can also notice that by taking only one term for the approximation of \( p_n \), 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 of the Cipolla development.

We hypothesize 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 don't know exactly. The best known formula for \( \pi(n) \) is that of \( \text{Li}(n) \). Riemann proposed a 2nd formula which seems much better at first sight but which was invalidated by Littlewood in 1914. This 2nd formula, called Riemann R or equivalently, 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 the Riemann R formula or the Gram series (converges faster) is excellent. But Littlewood has shown that after \( 10^9 \), the approximation drifts. As for the function \( \text{Li}(n) \), it behaves better at much larger scales, the first crossing being evaluated at around \( 10^{316} \). That is, \( \text{Li}(n) - \pi(n) = 0 \) around \( 1.397 \times 10^{316} \).
An interesting pattern appears in these calculations. If we analyze the error closely with the first term for formula (1) we quickly find a logarithmic type curve, then the question arises: what precision can we achieve if we use an approximation of the latter? For example, with

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

The ratio between the two is around 1 from the start and is around 0.999 when \( n = 10^{24} \). So, by calculating a curve of type \( a \log(n) + b \), we find a coefficient \( r^2 \) quite close to 1. The approximation is disappointing even if the coefficient is very high. Several formulas have been tested [11] to obtain the best accuracy.

The best approximation that has been found empirically is:

\[ p_n = -nW_{-1}\left(\frac{-e}{n}\right) - \frac{n \cdot c(n)}{W_0(n)} \]

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

What is happening is an effect of the Russian dolls, the matrioshkas. The rest with the first term is a curve which is roughly a straight line if you look from afar, the rest after 2 terms is still a ‘straight’ seen from afar but which is always of logarithmic type. And even with the final correction with \( c(n) \): the rest is still a logarithmic curve. All these curves seem equivalent but it is when we analyze closely the error that it varies.

By taking a sampling of the values of \( p_n \) between \( 10^2 \) à \( 10^{16} \)

| Step  | Number of values | Range                      |
|-------|------------------|----------------------------|
| \(10^2\) | 27117419         | 2711741900                 |
| \(10^3\) | 32082085         | 32082085000                |
| \(10^4\) | 45020269         | 450202690000               |
| \(10^5\) | 10000000         | \(10^{12}\)               |
| \(10^6\) | 4046531          | \(4.046531 \times 10^{12}\) |
| \(10^7\) | 5000000          | \(5 \times 10^{13}\)      |
| \(10^8\) | 454060           | \(4.54060 \times 10^{13}\) |
| \(10^9\) | 2200000          | \(2.2 \times 10^{15}\)    |
| \(10^{10}\) | 1112394         | \(1.112394 \times 10^{16}\) |
| \(10^{11}\) | 111239          | \(1.11239 \times 10^{16}\) |
| \(10^{12}\) | 54974           | \(5.4974 \times 10^{16}\) |
| \(10^{13}\) | 12317           | \(1.2317 \times 10^{17}\) |
| \(10^{14}\) | 2162            | \(2.162 \times 10^{17}\) |

We solve the equation for each \( n \) of the chosen table.

\[ -nW_{-1}\left(\frac{-e}{n}\right) - \frac{n \cdot x}{W_0(n)} - p_n = 0 \]
by the bisection method. The values are between 0.8 and 1. The logarithmic least squares curve is then calculated. The coefficient $r^2$ will indicate if the curve is right. The coefficients $a$ and $b$ are calculated 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 right. If $r^2$ is near 1 or 0, the curve follows a straight line very closely. The LLS (logarithmic least-squares) line is simply the log of the values that are aligned on a line.

For the range $100000 \ (10^5) \ldots 10^{12}$ we find:

$$0.00074741174603301665420395275429537 \ln(n) + 0.88596350453664534160747106131754$$

Once this formula is obtained, it remains to compare with the Cipolla formula. Here we will take 16 terms from Cipolla’s formula. In appendix, the 16 terms of development of Cipolla (program of B. Salvy).
Comparison with the range $100000 \ (10^5) \ldots \ 10^{12}$

$c(n) = 0.8859635045364534160747106131754 + 0.0007474117460330166542095275429537 \ ln(n)$

| Formula for $p_n$ | Cipolla (Salvy) | Lambert W |
|-------------------|----------------|-----------|
| Minimal gap       | 1624.006       | .0031723  |
| Maximal gap       | 7963203        | 3257663   |
| Average gap       | 3893600        | 617551    |

The formula with Lambert W is clearly more precise over the entire interval. In addition, some values are correct since the error is less than 0.5.

Comparison with the range $(10^{10}) \ldots 1.112394 \times 10^{16}$

$c(n) = 0.88281106024067112695415355478542 + 0.0008561837016404455723913114214399 \ln(n)$

| Formula for $p_n$ | Cipolla (Salvy) | Lambert W |
|-------------------|----------------|-----------|
| Minimal gap       | 640495         | 261       |
| Maximal gap       | 1004659553     | 513851652 |
| Average gap       | 4.64302e+08    | 1.25614e+08 |

More than 11 values reach an accuracy of 14 decimal digits.

**Results for $\pi(n)$**

For the calculation of $\pi(n)$, it suffices to reverse the formula for $p_n$. First of all, numerically it is very fast and above all very precise and even more precise than Li $(n)$.

So we pose that

$$\pi(n) \approx \left(-nW_{-1}\left(\frac{-e}{n}\right) - \frac{n c(n)}{W_0(n)}\right)^{-1}$$

That is, we solve the equation for a value $\pi(n)$ of the chosen table. With the same expression for $c(n)$.

Comparison with the range $100000 \ (10^5) \ldots 744000000000$

$c(n) = 0.8859635045364534160747106131754 + 0.0007474117460330166542095275429537 \ln(n)$

| Formula for $p_n$ | Li$(n)$ | Lambert W inverted | Gram or Riemann R |
|-------------------|---------|-------------------|-------------------|
| Minimal gap       | 37.8    | .00065655          | 0.0002465         |
| Maximal gap       | 53330.90| 19292.58           | 19205.24          |
| Average gap       | 24433.6 | 3659.52            | 3713.91           |
The formula with inverted Lambert W for the calculation of $\pi(n)$ is more precise than $Li(n)$ and happens to have a better average value with the scale $10^5$ but is not more precise for the scale $10^{10}$, but it is very close.
Appendice: formula of Cipolla with 16 terms.

\[ k \ln(k) \times (1 + \ln(\ln(k)) - 1) / \ln(k) + (\ln(\ln(k)) - 2) / \ln(k)^2 + \ldots \]

Formula: \( c^n \), \( n > 1 \). \( x \) is the nearest integer to \( x \). This formula gives 387 primes.
The primes are:

1000000001
31627769552311
1000000001478346301
3162777186062677745609
31000000221756159536498921059

8
The following constant c will generate 565 consecutive primes.

Formula : \( c^n \), \( n > 1 \). Where \( x \) is the nearest integer to \( x \).

Found on Feb. 13 2020.

c = 55237.075024967467154332147281528617412763740743534976625355574086706751108195952125636580713628080659, 
51610661608709511799686547852341008293602881791904714234505198633653577887369235086449112266789351, 
974241988869207456699900233385650655750943525653650761671263211343316696966739812219465780283859, 
54253548174652976500297130639457502661928886180484426067463553902812361363107119191473889991015, 
758877686624809743613134083754566222721736886865981592351899555262376530011834020688380214737993, 
939797139650327571991040629800467926396933233546984915866283575799080305405492649174215067261044294028, 
878866194815514557701042190356124366364824977111682579262512286348153978282106295507458352326384, 
50568410666741302350557805376823653323367365081804488468082231106639936566121734862691923929542, 
83385351381682790237630762318633822742846406336332106563349524469631120733692293212346, 
7554859262363953731419538287355327947055856146180559619268359623937960803779229620717693664434545, 
504858867437030200361307614178728627987883795230529966599650950718334389857600384328100591286183296999, 
6381848221022275811336061766849173325004816992862072902564895919259618737047150572523498197744990143307, 
8282998393367527423817899909684020618778888207968959061937839183362598945861060942516108142989397, 
2662673033224934166502891380338958468950325263512612112945767726260354767272444239004694006095661, 
46989787528799471968375270144971230447816901609620517455578138186367699381070202010816225448792465, 
45355406957089629974974060007448194996081616335395756901588828611325998121248257698067555979478728, 
...
References

[1] Encyclopedia of Integer Sequences, N.J.A. Sloane, Simon Plouffe, Academic Press, San Diego 1995.

[2] Mills, W. H. (1947). *A prime-representing function*, Bulletin of the American Mathematical Society 53 (6): 604, doi:10.1090/S0002-9904-1947-08849-2.

[3] E.M. Wright (1951). *A prime-representing function*. American Mathematical Monthly. 58 (9): 616–618. doi:10.2307/2306356. JSTOR 2306356.

[4] The OEIS, Online Encyclopedia of Integer Sequences, sequences: sequences A051021, A051254, A016104 and A323176.

[5] Wikipedia: formulas for primes (effective and non-effective formulas). https://en.wikipedia.org/wiki/Formula_for_primes

[6] Baillie Robert, The Wright’s fourth prime: https://arxiv.org/pdf/1705.09741.pdf

[7] Wikipedia: Le recuit simulé: https://fr.wikipedia.org/wiki/Recuit_simul%C3%A9

[8] Wikipedia: Simulated Annealing: https://en.wikipedia.org/wiki/Simulated_annealing

[9] László Tóth, A Variation on Mills-Like Prime-Representing Functions, ArXiv: https://arxiv.org/pdf/1801.08014.pdf

[10] Makoto Kamada Prime numbers of the form 7, 73, https://stdkmd.net/nrr/7/73333.htm#prime

[11] Plouffe, Simon: Pi, the primes and the Lambert W function, conference in July 2019, Montréal at the ACA 2019 (ETS), https://vixra.org/abs/1907.0108