The distribution of prime numbers: overview of n.ln(n)

Redoane D.*

E-mail: *redoane.daoudi@etu.univ-poitiers.fr

Abstract

The empirical formula giving the nth prime number p(n) is p(n) = n.ln(n) (from ROSSER (2)). Other studies have been performed (from DUSART for example (1)) in order to better estimate the nth prime number. Unfortunately these formulas don't work since there is a significant difference between the real nth prime number and the number given by the formulas. Here we propose a new model in which the difference is effectively reduced compared to the empirical formula. We discuss about the results and hypothesize that p(n) can be approximated with a constant defined in this work. As prime numbers are important to cryptography and other fields, a better knowledge of the distribution of prime numbers would be very useful. Further investigations are needed to understand the behavior of this constant and therefore to determine the nth prime number with a basic formula that could be used in both theoretical and practical research.

Preliminaries

We define the difference Δ as below:

$$\Delta = N - (n.ln(n)) \; n \in N* \; (1)$$

if N is the real nth prime number and $n.ln(n)$ the approximated nth prime number (see abstract).

We define ζ as below:

$$\zeta = (n.ln(n)) - \Delta \; n \in N* \; (2)$$

We define ϵ as below:

$$\epsilon = \tfrac{\zeta}{\Delta} \; \Delta \neq 0 \; (3)$$

The aim is to know Δ to find the real nth prime number. In fact Δ is the difference between the real nth prime number and the number given by the empirical formula. According to (2) we have:

$$\zeta = (n.ln(n)) - \Delta \; n \in N*$$

$$\zeta = (n.ln(n)) - \tfrac{\zeta}{\epsilon} \; \epsilon \neq 0$$

$$\zeta = \tfrac{\epsilon(n.ln(n)) - \zeta}{\epsilon} \; \epsilon \neq 0$$

$$\epsilon\zeta + \zeta = \epsilon(n.ln(n))$$

$$\zeta(\epsilon + 1) = \epsilon(n.ln(n))$$

$$\zeta = \tfrac{\epsilon(n.ln(n))}{\epsilon+1}$$

According to (3) we have:

$$\epsilon = \tfrac{\zeta}{\Delta} \; \Delta \neq 0$$

$$\Delta = \tfrac{\zeta}{\epsilon} \; \epsilon \neq 0$$

$$\Delta = \frac{\epsilon(n.ln(n))}{\epsilon^2+\epsilon} \ \epsilon \neq 0 \ (4)$$

Finally the real nth prime number is given by the following formula:

$$N = (n.ln(n)) + \Delta \ n \in N*$$

$$N = (n.ln(n))\frac{2+\epsilon}{1+\epsilon} \ \epsilon \neq -1 \ (5)$$

Consequently, we must to know ϵ to find Δ and the real nth prime number. In this work we discuss about the value of ϵ that is associated with four intervals and two values for n (see Methods and results). Although there are variations for ϵ we define it as a constant because ϵ seems to show small variations (its value is -0.14092488 for n=2 and 7.271015283 for n = 10^6).

Methods and results

Methods

In this work we use four intervals and two values for n as described below:

$$n \in [2, 200] \quad n \in [1000, 1195] \quad n \in [1800, 1999] \quad n \in [2600, 2799] \quad n = 100000 \ n = 10^6$$

Note that the gap between each interval is about 600 (except for the first).

By using Microsoft Excel 2016 and a list of known prime numbers we define, for each n of each interval, the corresponding prime number. For example the prime number N corresponding to n=2 is 3. If n=5, the prime number N that is associated is 11.

Then, for each nth prime number, we calculate n.ln(n) that is the approximated nth prime number, Δ, ζ and ϵ. Finally, for each interval, we determine the average value for ϵ. We define ϵ' as the average value for ϵ. Our results suggest that ϵ' shows small variations even if n is comprised between 2 and 10^6.

Finally, we calculate for each n of each interval, the corresponding nth prime number given by our method with the formula (5). We define this number as N' and we determine the difference between N' and the real nth prime number as Δ'. Our results show that the average value for Δ' is significantly smaller than the average value for Δ.

Results

Note the corresponding values for ϵ' for each interval:

$$n \in [2.200] \quad \epsilon' = 5.1504379$$

$$n \in [1000.1195] \quad \epsilon' = 5.81488845$$

$$n \in [1800.1999] \quad \epsilon' = 6.09402715$$

$$n \in [2600.2799] \quad \epsilon' = 6.19761627$$

$$n = 100000 \ \epsilon = 6.757176176$$

$$n = 10^6 \ \epsilon = 7.271015283$$

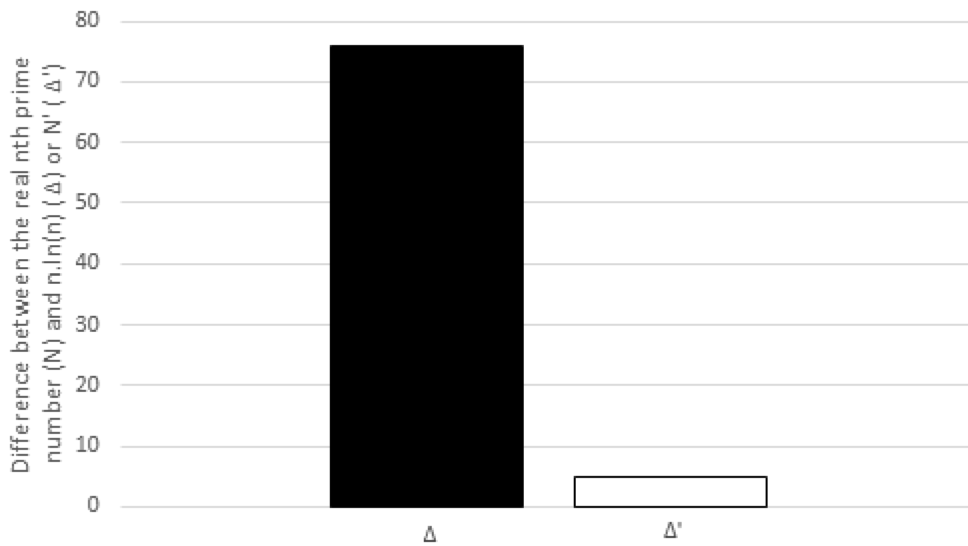

Figure 1: Δ' (4.87377461) is significantly smaller than Δ (75.819752) (average value). Results for the interval $[2, 200]$

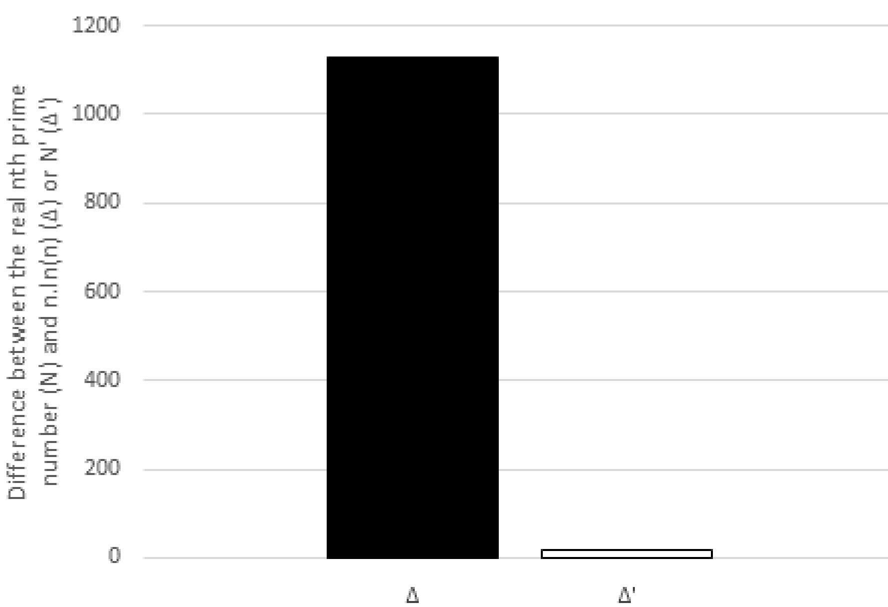

Figure 2: Δ' (15.0668754) is significantly smaller than Δ (1127.19363) (average value). Results for the interval $[1000, 1195]$

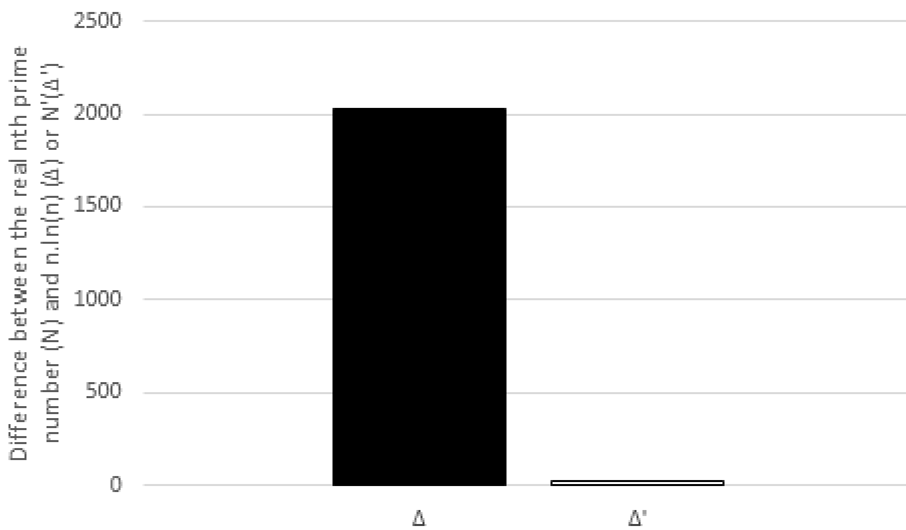

Figure 3: Δ' (19.3845647) is significantly smaller than Δ (2022.39968) (average value). Results for the interval $[1800, 1999]$

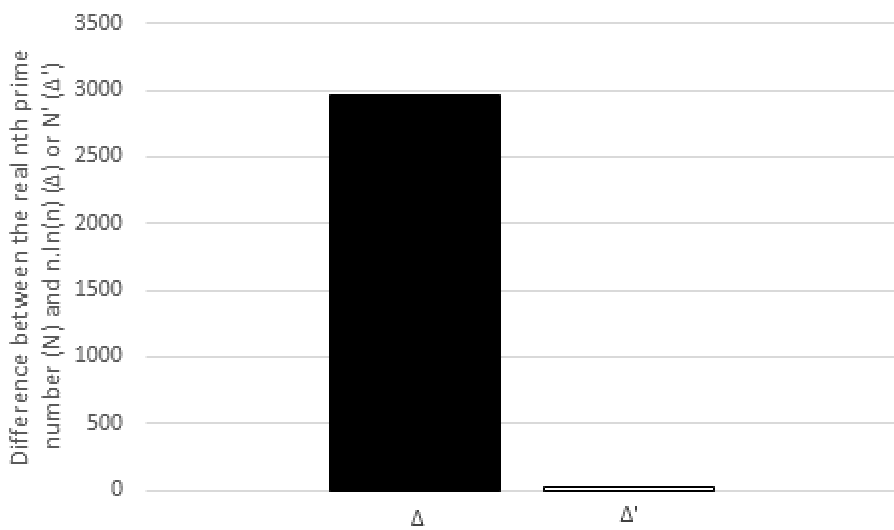

Figure 4: Δ' (29.0238221) is significantly smaller than Δ (2963.99395) (average value). Results for the interval $[2600, 2799]$

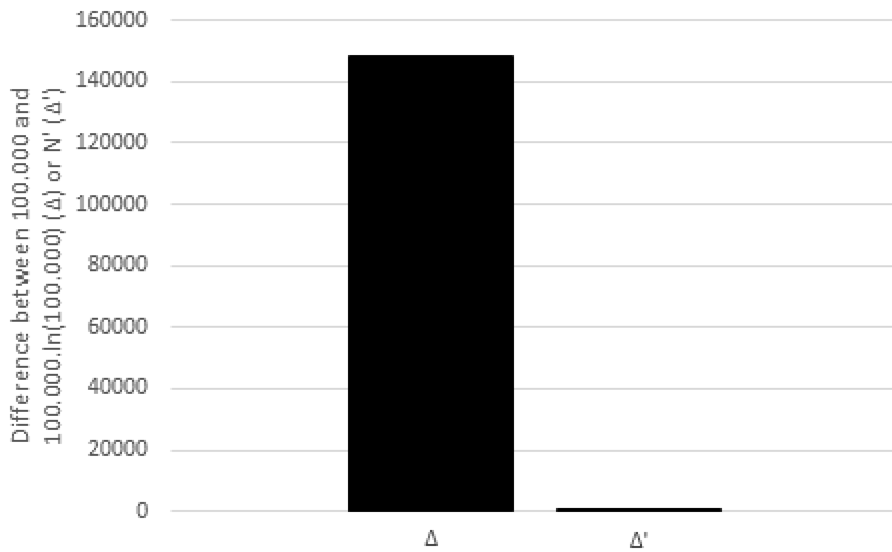

Figure 5: $\Delta' = 0$ is significantly smaller than Δ (148416). Results for the value n=100000

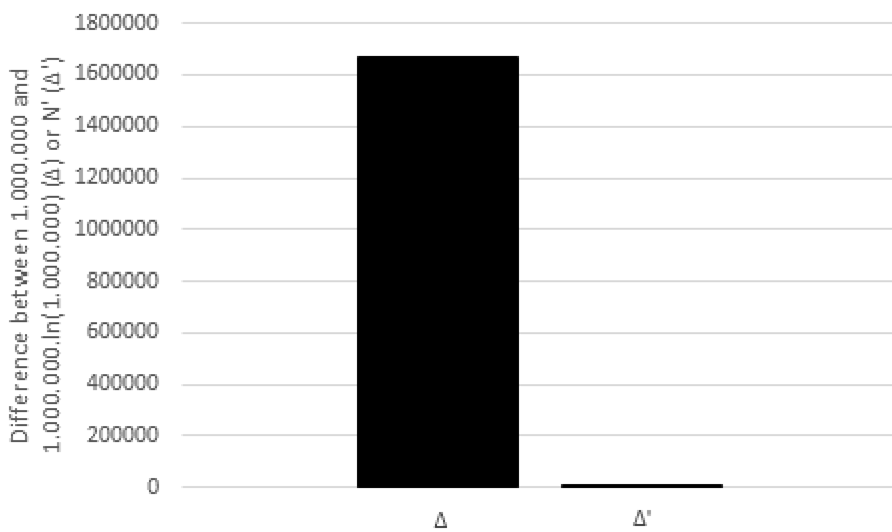

Figure 6: $\Delta' = 0$ is significantly smaller than Δ (1670352). Results for the value $n = 10^6$

Discussion

We have established a relationship between the real nth prime number and a constant ϵ. The use of our method shows results much more effective compared to the empirical formula. For certain values of n, we found the real nth prime number (i.e $\Delta' = 0$). For example, if n=100000 the real nth prime number that is associated is 1299709. The best value for ϵ for n=100000 is 6.757176176. With this value, we find the real prime number ($\Delta' = 0$). Unfortunately the best value for ϵ is found only if we know the real nth prime number. If the real nth prime number is unknown, the best value for ϵ remains to be elucidated and the best value is required to find the real nth prime number. If we know ϵ it is possible to determine the precise nth prime number ($\Delta' = 0$ since there is a link between Δ and ϵ (see (4)).

Interestingly ϵ seems to show very small variations whereas n is comprised between 2 and 10^6. For this reason, we hypothesize that ϵ could be a very good constant in order to find a formula establishing a more precise link with the nth prime number. Nevertheless small variations for ϵ are responsible for greater variations for Δ'.

ϵ seems to exhibit very small variations while n allows high values. However ϵ shows high variations between the first and the second interval (5.81488845-5.1504379 = 0.66445055 for a difference of n=800 between the two intervals). ϵ shows smaller variations between the last interval and 100000 (6.757176176-6.19761627=0.559559906) and between 100000 and 10^6 (7.271015283-6.757176176=0.513839107). These results suggest that ϵ is not defined by the length of the gap between two intervals because in this case ϵ would be much more greater than 7.271015283 for n=10^6. Furthermore, these results suggest that

$$\epsilon \underset{n \to +\infty}{\longrightarrow} k$$

if $k \in R$ and $k > 7.271015283$

If this is the case, ϵ is a constant when $n \to +\infty$ and it will be easier to find a formula to determine nth prime numbers when n tends to infinity. But the most likely hypothesis is that ϵ doesn't converge when n tends to infinity and it will be more difficult to find a formula to determine nth prime numbers.

Interestingly we notice that if we consider a specific nth prime number (in the third interval of this study for example), the value of ϵ of the next or the previous prime number is very close to the value of ϵ for the specific nth prime number considered. This is very useful because we could imagine a formula giving the next prime number if the previous is known. However this is an observation with exceptions and there is no evidence when $n \to +\infty$. Moreover, even if the value of ϵ of a next prime number seems to be close to the value of ϵ of a previous prime number, there are unpredictable variations and a little mistake for the value of ϵ results in a big error in the determination of the nth prime number. Note that the value of a specific ϵ can be smaller than the value of ϵ of a previous prime number. Moreover, for certain values of n we observe exceptions, for example for n=30 (the real nth prime number corresponding is 113) we found $\epsilon = 8.30638366$, a value that is greater than the value of ϵ that is associated with n=10^6 (7.271015283). These observations suggest there are several exceptions for the value of ϵ. However the average value for ϵ (ϵ') seems to be greater than the value for ϵ' that is associated with smaller values for n. Therefore the exact relationship between n and ϵ is unknown and may not exist. Thus, the behavior of ϵ remains to be elucidated (see below).

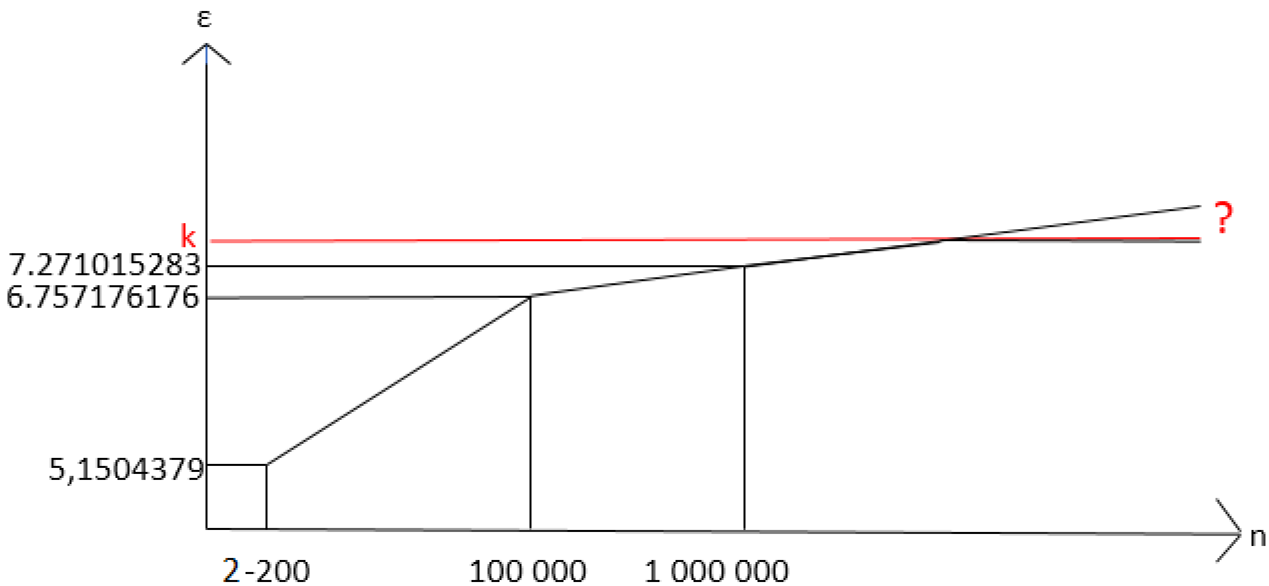

Figure 7: The behavior of ϵ remains to be elucidated when $n > 10^6$. ϵ may converge to k when $n \to +\infty$ or not.

Conjecture

For $n > 2$ ϵ seems to be comprised between 0 and $+\infty$ (if ϵ doesn't converge when $n \to +\infty$).

So we hypothesize that:

$$N = (n.ln(n))\frac{2+\epsilon}{1+\epsilon}$$

$$N = (n.ln(n))p \text{ with } 1 < p < 2,\ p = \frac{2+\epsilon}{1+\epsilon} \text{ and } n > 2$$

In fact -1 (for n=1) and -0.14092488 (for n=2) seem to be the only negative values for ϵ. When $n \to +\infty$ and if ϵ doesn't converge we have $p \to 1$ because:

$$\lim_{\epsilon \to +\infty} \frac{2+\epsilon}{1+\epsilon} = 1$$

But p is not exactly 1 and a small variation of p (for example 1.00000005 instead of 1.00000004) is responsible for a big error in the determination of big nth prime numbers.

Now, when n is smaller ϵ is also smaller and we have:

$$\lim_{\epsilon \to 0} \tfrac{2+\epsilon}{1+\epsilon} = 2$$

This is the reason for which we hypothesized that $1 < p < 2$. However this is a conjecture because it is not sure if -1 and -0.14092488 are the only negative values for ϵ, even if it is very likely.

Conclusion

We established a model in which we were able to find the real nth prime number by using a new constant called ϵ. When this constant is known precisely, the nth prime number is known precisely. If this constant is imprecise, the nth prime number will be imprecise. Surprinsingly we found that ϵ would be comprised between -0.14092488 (for n=2) and 7.271015283 (for n=10^6), even if n undergoes high variations (from 2 to 10^6). But small variations in ϵ result in an imprecise formula and there are several exceptions with values of ϵ that are greater than 7.271015283. For this reason, further investigations are needed to understand the behavior of ϵ and to establish a potential relationship between n and ϵ. For example a new work is necessary to study all prime numbers between 3 and 10^6 (not shown in this study because we worked with only four intervals and two values for n).

In all cases, even if ϵ isn't known precisely, our formula seems to be more precise than the empirical formula n.ln(n).

Tools

Statistics. Statistics were performed using Microsoft Excel 2016.

The list of prime numbers used in this study. http://compoasso.free.fr/primelistweb/page/prime/liste_ online.php

References

1. PIERRE DUSART, The k^{th} prime is greater than $k(lnk + lnlnk - 1)$ for $k \geq 2$, Math. Comp. 68 (1999), 411-415

2. J.B. ROSSER, The n^{th} prime is greater than $n.log(n)$, Proc. London Math. Soc. (2) 45 (1939), 21-44