

Note on the Cardinality Difference between Primes and Twin Primes and its Impact on Function $x/\ln(x)$ in Prime Number Theorem

Sandor Kristyán

*Research Centre for Natural Sciences, Hungarian Academy of Sciences,
Magyar tudósok körútja 2, Budapest H-1117, Hungary*

Corresponding author: kristyan.sandor@tk.mta.hu

Keywords

prime number theorem,
analysis of $x/\ln(x)$ in relation to twin primes and PNT

Abstract

We provide a simple explanation and test that the emblematic function $x/\ln(x)$ in prime number theorem estimates roughly the prime-counting as 1 for each isolated primes and 1 for each twin primes, and not 2 for twins as it should be. By this reason, for a more proper counting the function $(x/\ln(x))[1+\ln(x)^{-1}]$ should replace $x/\ln(x)$ in the prime number theorem.

Discussion

The prime number theorem (PNT) describes the asymptotic distribution of the prime numbers in the set of positive integers. The first distribution found is

$$\pi(x) \approx x/\ln(x) \tag{1}$$

by C.F. Gauss in 1792, and later the better $\pi(x) \approx \text{Li}(x) \equiv \int_{(2,x)} (\ln(t))^{-1} dt$ by P.G.L. Dirichlet in 1838, both using the natural logarithm. The $\pi(x)$ is the prime-counting function that gives the number of primes $\leq x$ for any real number x , e.g. $\pi(\pi=3.14\dots) = 2$, $\pi(10) = 4$ or $\pi(10^6) = 78498$, etc.. The PNT states that $x/\ln(x)$ is a good approximation to $\pi(x)$ in the sense that the limit of the quotient $\pi(x)/(x/\ln(x)) \rightarrow 1$ as $x \rightarrow \infty$, i.e. Eq.1 is true as an asymptotic law of distribution of prime numbers, so the more precise notation is $\pi(x) \sim x/\ln(x)$. The error $(\pi(x) - x/\ln(x))$ by $x/\ln(x)$ is larger and larger, but “error”/ $\pi(x) \rightarrow 0$ as $x \rightarrow \infty$, (see curves E and H on Figure 1). However, Eq.1 is also valid as approximation for “larger” x , so it is tabulated and plotted continuously in the literature, - that is the target we comment in this work. The reason that approximation (\approx) vs. asymptotic behavior (\sim) in Eq.1 must be emphasized is that its “error” $\rightarrow \infty$ also as $x \rightarrow \infty$ (see table below). For large enough N , the probability that a random integer not greater than N is prime is very close to the integrand $(\ln(N))^{-1}$ in $\text{Li}(x)$, see also ref.[1] wherein discrete valued function is introduced for distribution instead of the continuous valued $\ln(x)$. It follows via elementary probability theory that this kind of probability for a twin prime is about the product $(\ln(N))^{-2}$, so similarly, $\pi_2(x) \approx \int_{(2,x)} (\ln(t))^{-2} dt$, where $\pi_2(x)$ denotes the number of twin primes p and $p+2$ such that $p \leq x$, e.g. $\pi_2(10) = 2$ being (2,3) and (5,7) the only two twin primes under 10 or $\pi_2(10^6) = 8169$, etc.. (More precisely, $\pi_2(x) \approx 2\Pi_2 \int_{(2,x)} (\ln(t))^{-2} dt$, where $\Pi_2 = 0.66016\dots$ is the twin primes constant which has more importance in the limit behavior.) Additionally, famous ancient theorem states $\pi(\infty) = \infty$, and yet unproved conjecture states $\pi_2(\infty) = \infty$, as well as for the n^{th} prime the obvious $\pi(p_n) = n$ provides the very first approximation such as $p_n \approx n \ln(n)$. Primes which are not twin primes are called isolated primes. Technically, $\pi(x) = \pi_{\text{isolated}}(x) + 2\pi_2(x)$, so the factor 2 is necessary. These are the number one behaviors which can be found in textbooks, the formulations are simple, at least, but not the theory behind. (First proofs of Eq.1 have appeared in 1896 only, using Riemann zeta function, but proofs were published even after 100 years also, when easier ways were found.) One can also find these even in internet (such as Wikipedia or math in Wolfram) along with large lists of primes and twin primes.

We do not even try to review the literature how these expressions (i.e. for $|\pi(x) - \text{Li}(x)|$ and $|\pi(x) - x/\ln(x)|$) are developed, becoming more and more complex even in the formulations itself. We mention only one old and easy improvement: e.g. the $x/(\ln(x)-c)$ with $c \approx 1$ strongly improves $x/\ln(x)$. We emphasize that in regard to PNM, the two most characteristic functions are the $x/\ln(x)$ and $\text{Li}(x)$, no question about it, the former is more explicit what mathematician desires, but the latter is more accurate. For our talk *in medias res* we say that error of $x/\ln(x)$ for estimating $\pi(x)$ is about “error” $\approx \pi_2(x)$, because probability $1/\ln(N)$ at larger N (very visible from $N \approx 1000$, see table below) sees both, isolated and twins as 1 entity, but it should be 2 for twin in counting for $\pi(x)$, that is, $x/\ln(x)$ counts $\pi_{\text{isolated}}(x) + \pi_2(x)$ and not the proper $\pi_{\text{isolated}}(x) + 2\pi_2(x)$, this will be analyzed and supported mathematically below.

We start with one simple analytic relationship which is worth to mention immediately between Eq.1 and $\text{Li}(x)$. They elementary relation comes from $d(x/\ln(x))dx = (\ln(x)-1)/(\ln(x))^2 \sim 1/\ln(x)$ for large x , i.e. the integrand in $\text{Li}(x)$ is recovered asymptotically. Although $\text{Li}(x)$ cannot be integrated analytically, one can try with $(x+g(x))/\ln(x)$ to find which kind of $g(x)$ yields $1/\ln(x)$ after derivation. It yields a first order, inhomogeneous, ordinary differential equation with function coefficients, and the standard way to the solution yields $g(x) = (\ln(x)) \int (\ln(x))^{-2} dx$, (not in contradiction that no analytic primitive function). The back-check provides the idea to consider $(\ln(x))^{-1} = d(x/\ln(x))/dx + (\ln(x))^{-2}$: evaluation of the derivative in it yields triviality, but its integration (web page wolfram recognizes this trick) yields the integral evaluation for a half way and a key equation for our talk as

$$\int (\ln(x))^{-1} dx = x/\ln(x) + \int (\ln(x))^{-2} dx. \quad (2)$$

To the compact and exact Eq.2 the known asymptotic expansion should also be mentioned: $\text{Li}(x) \sim (x/\ln(x)) [1 + (\ln(x))^{-1} + 2(\ln(x))^{-2} + \dots + k!(\ln(x))^{-k} + \dots]$ for comparison. Eq.2 is the key expression for us, because its consequence is

$$\pi(x) - \pi_2(x) \approx x/\ln(x) \quad (3)$$

as a very simple and powerful improvement over Eq.1, because it explains where the main part of “error” comes from in $x/\ln(x)$. It means that $x/\ln(x)$ approximates the number of primes minus the number twin primes rather than the number of primes ($\leq x$). Eq.2 explains the well known fact why $\text{Li}(x)$ approximates the $\pi(x)$ much better than the $x/\ln(x)$ does. To the knowledge and experience of the author, only e.g. plots for ratios $\pi(x)/(x/\ln(x))$ vs. $\pi(x)/\text{Li}(x)$ are exhibited, or the three functions ($\pi(x)$, $x/\ln(x)$, $\text{Li}(x)$) together and similar kind of statistics, but Eq.2 which is immediately responsible for the difference is not recognized in this respect, although it should be. Furthermore, dividing Eq.3 by $x/\ln(x)$ yields that $(\pi(x) - \pi_2(x))/(x/\ln(x)) \rightarrow 1$ via this limit behavior of $\pi(x)$, and as a consequence $\pi_2(x)/(x/\ln(x)) \rightarrow 0$ as $x \rightarrow \infty$.

Definite integral in Eq.2 on interval $(2, x)$ with crude $\pi_2(x) \approx \int_{(2,x)} (\ln(t))^{-2} dt$ and the asymptotic expansion for the more accurate $\pi(x) \approx \text{Li}(x)$ along with stopping at the second term, yields that $\pi_2(x) \approx x/(\ln(x))^2$, and inserting the latter into Eq.3 yields

$$\pi(x) \approx x/\ln(x) + x/(\ln(x))^2, \quad (4)$$

a formula known in boundary analysis for $\pi(x)$. The error estimation/improvement by Eq.4 over Eq.1 comes from the form $\pi(x) \approx (x/\ln(x)) [1 + (\ln(x))^{-1}]$, i.e. error $\approx 100\% (\ln(x))^{-1}$. For example, at $x = 10^{16}$ it is only 2.7%, but caution must be, since it is large, because $\pi(x)$ as well as $x/\ln(x)$ rapidly grow with x (Figure 1.a). Beside the analytical support by Eq.2, the probability approach is as follows: $1/\ln(N)$ gives the probability to find an isolated prime or twin prime, but at larger N it misses the greater (or smaller) prime in twins $(p, p+2)$ at that N . However, $1/(\ln(N))^2$ gives the probability to find a twin, so the sum $1/\ln(N) + 1/(\ln(N))^2$ count more properly (re-correct) to find the number of all primes approximately. For example, at $N = 10^6$ the $\pi = 78498$ and $\pi_2 = 8169$, i.e. $x/\ln(x)$ mistakenly counts one instead of two for the 8169 twins, and indeed, $10^6/\ln(10^6) = 72382.4$ counts for $\pi - \pi_2 = 78498 - 8169 = 70329$, while $10^6/\ln(10^6) + 10^6/(\ln(10^6))^2 = 77621.6$ counts for 78498. This is the way how Eq.4 improves for any, but mainly for large x . There are other “close packages” of primes, e.g. “cousin primes” $(p, p+4)$ or “sexy primes” doublets, triplets and quadruplets $(p, p+6, p+12, p+18)$ even for very large primes p , but they have less importance than twins. However, no higher multiplicity than 2 with difference 2, i.e. no higher multiplicity than twins $(p, p+2)$, since one among $(N, N+2, N+4)$ can always be divided by 3; exception is at the beginning: the $(2, 3, 5, 7)$ is a quintet. The superiority of $\text{Li}(x)$ over $x/\ln(x)$ may come from its ability to count for these close packages better, nicely visible by the higher power terms in its asymptotic expansion via $k!/(\ln(x))^k$.

We stop at this point, but a simple fit for parameter “a” in

$$\pi(x) \approx x/\ln(x) + ax/(\ln(x))^2 \quad (5)$$

strongly improves Eq.4 as an empirical scaling. Curves J, K and L on Figure 1.b below provides a little scan with computer search in the range x in $[10, 10^{16}]$ yielding an optimal value $a = 1.0595$, i.e. it does not

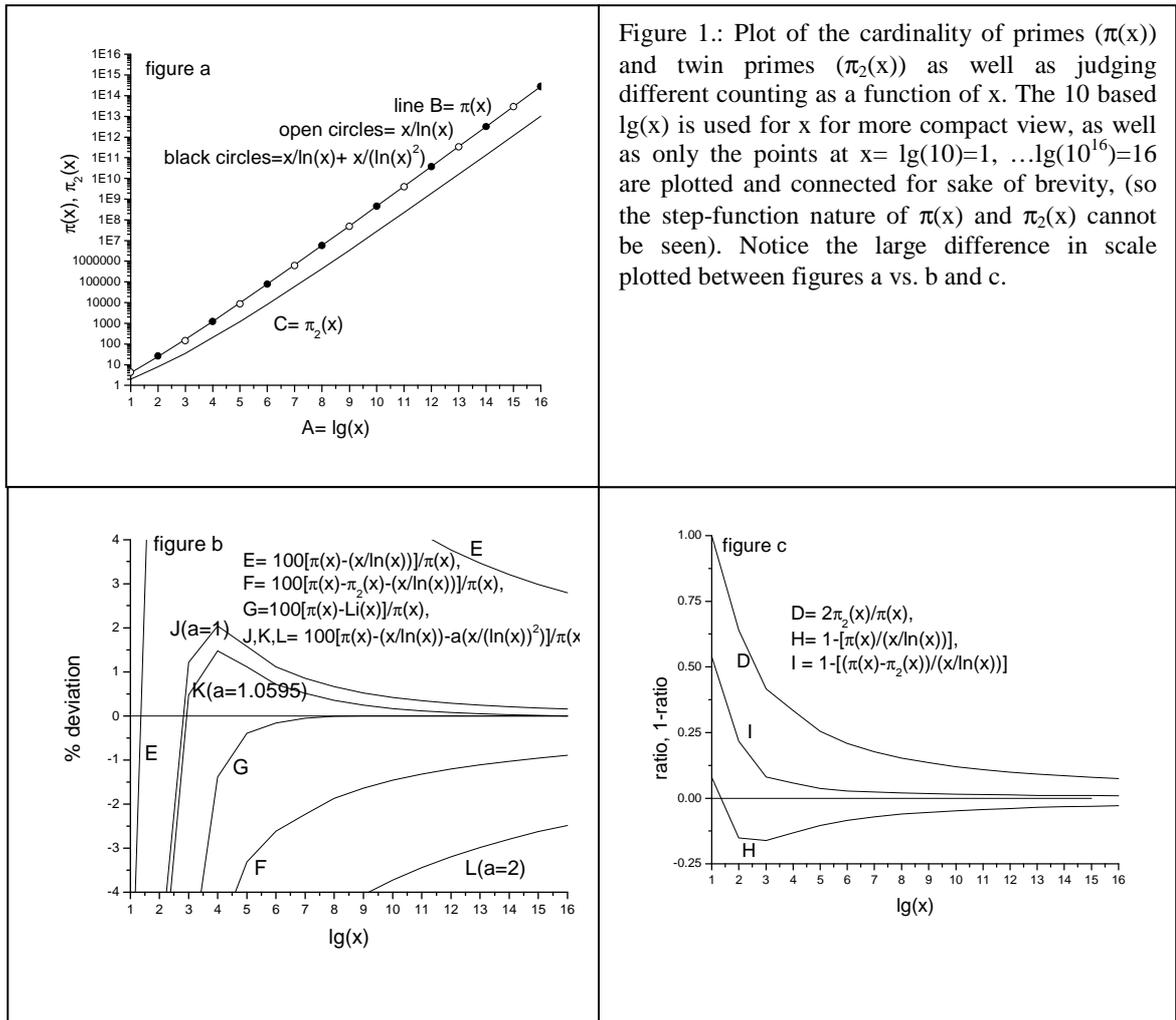
deviate much from unity, as well as in agreement with twin primes constant as $1 < 2\Pi_2 < 2$. The parameter “a” is very sensitive in the short interval [1,2] (compare curves J and K, as well as L exhibits sign change), and we note that scaling device in Eq.5 is just a very simple way to improve Eq.4. More importantly, this optimal ≈ 1 greatly supports the idea behind Eq.4. To go further, a more subtle theory needed, but what is important now is that Eq.4 is a strong improvement over Eq.1. We do not review and seek more accurate forms than Eq.4, instead we are explaining the meaning of terms in Eqs.1 and 4. In relation to PNT, the limit behavior for Eq.4 is $(x/\ln(x) + x/(\ln(x))^2)/(x/\ln(x)) = 1 + 1/(\ln(x)) \rightarrow 1$ as $x \rightarrow \infty$, which includes the more important $\pi(x)/(x/\ln(x) + x/(\ln(x))^2) \rightarrow 1$ as $x \rightarrow \infty$, because by this: $\pi(x) \sim (x/\ln(x) + x/(\ln(x))^2)$.

Eq.2 is an exact equation, but Eqs.1 and 3-5 are only approximations. In this way we tabulate and plot $\pi(x)$, $\pi_2(x)$, $\text{Li}(x)$, $x/\ln(x)$ and $x/(\ln(x))^2$ in different combinations for deviations and ratios as functions of x for the range $[10, 10^{16}]$ to show that Eqs.3-5 are superior over Eq.1. The tabulation is

$\lg(x)$	$\pi(x)$	$\pi(x) - [x/\ln(x)]$	$\pi(x) - [(x/\ln(x)) + (x/(\ln(x))^2)]$
1	4	-0.34	-2.23
2	25	3.29	-1.43
3	168	23.24	2.28
4	1229	143.26	25.38
5	9592	906.11	151.66
6	78498	6115.59	876.37
16	279238341033925 $\approx 279.2 * d12$	7.804 * d12	0.437 * d12

showing that Eq.4 nicely improves Eq.1. For visualization the plots: Figure 1.a reveal that there are not only many primes, but many twins, and most importantly the genius behavior of $x/\ln(x)$ is visible, but at this resolution eye cannot distinguish curve B from circles, i.e. the effect between Eqs.1 and 4 cannot be seen. Curve D on Figure 1.c shows the cardinality ratio between twins and primes. Curves on Figure 1.b exhibit error percentages: Comparison of E and F shows that Eq.3 is superior over Eq.1, even the sign changes, as well as J, K and L exhibit the above mentioned scaling for Eq.4 via Eq.5, and J vs. F reveal that Eq.4 is even better than Eq.3, probably via error cancellation between the two terms in Eq.4; but G proves that $\text{Li}(x)$ does best for limit behavior among the functions considered. Figure 1.c exhibits the ratios from unity for the asymptotic behavior, and comparing curves H and I reveals that the convergence is faster by Eq.3 than in Eq.1 with opposite sign, that is convergence from above vs. below. In summary, curves E and J on Figure 1.b reveals that both, Eq.4 and Eq.1 perform correct asymptotic behavior, but Eq.4 is more accurate (in “ \approx ”) and faster (in “ \sim ”) via more precious counting.

We know that the block functions mentioned above are well known in the theory of primes, but we want to emphasize that Eq.2 from calculus is not recognized enough as important one in the theory of primes how it provides immediate improvement, as well as - as a consequence of Eq.2 - the meaning of terms $(x/\ln(x))$ and $x/(\ln(x))^2$ has been made clearer in Eqs.3-5 to understand how $\pi(x)$ is estimated in its first well known and emblematic step in Eq.1. We have shown that Eq.3 is more accurate than Eq.1, or with substituting $\pi_2(x) = (\pi(x) - \pi_{\text{isolated}}(x))/2$, Eq.3 reads as $\pi(x) + \pi_{\text{isolated}}(x) \approx 2x/\ln(x)$. Eq.3 tells what $x/\ln(x)$ counts in fact, and Eq.4 provides a better estimation for $\pi(x)$ than Eq.1. As a final word, the more accurate Eq.4 should replace the simpler Eq.1 in PNT for both relations: \approx and \sim once and forever, the simplicity of Eq.4 deserves this in an emblematic theorem like PNT. (In fact the $\pi(x) \approx \text{Li}(x)$ is the best among discussed, but its handicap is the non-existing (or lack of) analytical primitive function.)



ACKNOWLEDGMENTS

Financial and emotional support for this research from OTKA-K 2015-115733 and 2016-119358 are kindly acknowledged.

REFERENCES

- 1.: Sandor Kristyan: On the statistical distribution of prime numbers: A view from where the distribution of prime numbers are not erratic, AIP Conference Proceedings (on ICNAAM 2016) **1863**, 560013 (2017); <http://doi.org/10.1063/1.4992696>