# Two Elementary Connections Between the Number of Primes and the Logarithmic Integral

Sometime around 1792, at age 15 or 16, Gauss speculated that the logarithmic integral was a good approximation for the number of primes less than *n*... and it is! But that was an empirically-based theory on his part.

Then, in 1859, Bernhard Riemann actually published an explicit formula for the number of primes less than *n*. That formula showed the precise difference between the number of primes and the logarithmic integral, encoded in the zeros of the Riemann zeta function. That same paper also introduced the now very famous Riemann Hypothesis regarding those zeros. One immediate consequence of the hypothesis being true would be that the difference between the number of primes and the log integral is provably smaller than if the hypothesis is false.

It's fascinating, profound stuff. Unfortunately, it's also really challenging stuff. If you're interested in the distribution of the primes, the concepts involved in understanding Riemann's formula require some pretty heavy mental lifting, particularly if you're not an expert at complex analysis.

It might come as a surprise to you that there are other ways of connecting the prime counting function and the logarithmic integral, ways that don't require any knowledge of complex analysis whatsoever.

After reviewing Riemann's explicit formula connecting the Riemann prime counting function to the logarithmic integral, this article covers two such elementary connections. This isn't a formal math paper; we won't prove anything. Rather, we'll take a more casual, experimental, empirical approach. Interspersed with the identities will be Mathematica code to encourage tinkering with the ideas.

## The Number of Primes and the Logarithmic Integral

We want to talk about the difference between the number of primes and the logarithmic integral. So we need to define those first.

Mathematicians usually use $\pi(n)$ to mean the count of numbers <= n that are prime, but it's often not the most natural prime counting function. Instead we'll use a more natural one, the Riemann prime counting function:

$$\Pi(n) = \sum_{k=1}^{\lfloor \log_2 n \rfloor} \frac{1}{k} \pi(n^{\frac{1}{k}})$$

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[ n^(1/k)]/k, {k,1,Log[2,n]}]
Table[{ n, referenceRiemanPrimeCount[ n ]}, { n, 1, 100 }] // TableForm
PrimePiAlt[n_]:=Sum[MoebiusMu[k] referenceRiemanPrimeCount[ n^(1/k)]/k, {k,1,Log[2,n]}]
Table[{ n, PrimePi[ n ], PrimePiAlt[n] }, { n, 1, 100 }] // TableForm
```

Recovering $\pi(n)$ from $\Pi(n)$ is trivial, as shown by PrimePiAlt above, so we won't mention it again.

So, we have our prime counting function. Now we need the approximation that Gauss suggested, the logarithmic integral. It can be defined something like this:

$$li(x) = \int_0^x \frac{dt}{\log t}$$

```
Table[{n,N[LogIntegral[n]], Integrate[ 1./Log[x],{x,0,n}, PrincipalValue->True]},{n,2,100}]//TableForm
```

N

$$li(x) - \log\log x - \gamma = \int_1^x \frac{1}{\log t} - \frac{1}{t \log t} dt$$

```
Table[{x,N[LogIntegral[x]-Log[Log[x]]-EulerGamma], N@Integrate[ 1/Log[t]-1/(t Log[t]),{t,1,x}]},{x,2,100}]//TableForm
```

Now the literal million dollar question: how good is the approximation?

Well, if you have a copy of Mathematica handy, try graphing these functions:

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[ n^(1/k)]/k, {k,1,Log[2,n]}]
DiscretePlot[ {referenceRiemanPrimeCount[n],LogIntegral[n]},{n,1,1000}]
DiscretePlot[ {referenceRiemanPrimeCount[n],LogIntegral[n]},{n,1,100000, 100}]
DiscretePlot[ {referenceRiemanPrimeCount[n],LogIntegral[n]},{n,1,1000000000,1000000}]
```

Looks like a tight approximation, doesn't it? Let's look more closely at the difference between the two:

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[ n^(1/k)]/k, {k,1,Log[2,n]}]
DiscretePlot[ {referenceRiemanPrimeCount[n]-LogIntegral[n]},{n,1,10000,10}]
```

You might wonder, looking at these pictures, how big the difference can be.

That's where the Riemann hypothesis comes in. If the Riemann Hypothesis is true, the difference between $\Pi(n)$ from $li(n)$ for *any* n, no matter how gargantuan, will satisfy

$$|\Pi(n)-li(n)| < \frac{1}{8\pi}\sqrt{n}\log n \quad \text{for } n \geq 59$$

So in this graph

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[n^(1/k)]/k,{k,1,Log[2,n]}]
DiscretePlot[ {referenceRiemanPrimeCount[n]-LogIntegral[n], n^(1/2) Log[n]/(8 Pi ), -n^(1/2) Log[n]/(8 Pi )},{n,1,1000000000,1000000}]
```

the erratic line in the middle will never, ever, ever cross the two smooth curves bounding it... if the Riemann Hypothesis is true. And vice versa, too – if there were some other way to prove $|\Pi(n)-li(n)| < \frac{1}{8\pi}\sqrt{n}\log n \quad \text{for } n \geq 59$ , that would prove the Hypothesis.

Got it?

So that provides some pretty good gut-level evidence that there must be some connection between the two functions. But it really doesn't give any mathematical insight, of course. If you're a math-y sort of person, this feels like the start of a conversation, certainly not the end of it.

---

# Riemann's Explicit Formula

Here is Riemann's famous Explicit Formula.

$$\alpha(n) = \frac{1}{2}\frac{\Lambda(n)}{\log n} - \log 2 + \int_n^\infty \frac{dt}{t(t^2-1)\log t}$$

$$\Pi(n) = li(n) - \sum_\rho li(n^\rho) + \alpha(n)$$

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[n^(1/j)]/j,{j,1,Log[2,n]}]
al[n_]:=MangoldtLambda[n]/2/Log[n]+NIntegrate[1/((t^3-t) Log[t]),{t,n,Infinity}]-Log[2]
RieExplicitForumla[n_,t_]:=LogIntegral[n]-2Re@Sum[ExpIntegralEi[N@ZetaZero[k] Log[n]],{k,1,t}]+al[n]
Table[{n,N[referenceRiemanPrimeCount[n]],RieExplicitForumla[n,500]},{n,2,100}]//TableForm
```

Before you even try to make sense of this, note that you can ignore the function $\alpha(n)$ . For any $n>1$ , $\alpha(n)$ is bounded by $-1<\alpha(n)<0$ , so its contribution to the distribution of primes is nearly non-existent.

is Riemann's Explicit Formula saying? Well, it essentially expresses the Riemann Prime Counting function $\Pi(n)$ as the smooth logarithmic integral $li(n)$ plus this other sum, $\sum_\rho li(n^\rho)$ .

Let's talk about that sum a bit. In Mathematica, we calculate it as `2Re[ Sum[ ExpIntegralEi[ N@ZetaZero[ k ] Log[ x ]], { k, 1, t }]]` . ( $\rho$ here is shorthand for the zeros of the Riemann zeta function) Why's the Mathematica code differ from the formula, you might very well wonder?

Well, there are an infinite number of zeta zeros, each with successively diminishing impact on our final sum. Given that we would like our computation to stop before the heat death of the universe, we approximate using a limited number of zeros.

Also, every zeta zero comes with a complex conjugate – an almost identical zero who's imaginary part has its positive/negative signed flipped. If we used both, we'd find the resulting imaginary parts would cancel out, leaving us with 2 times the real part of using just one. So we only use on zero of the pair, toss its imaginary result, and multiply by 2.

Finally, our definition (and Mathematica's definition) of the logarithmic integral is only valid here for real numbers. So we calculate the logarithmic integral of $n^\rho$ , a complex number, through other means.

Try playing around with the zeros a bit, perhaps using a command like`Table[ {N[ZetaZero[x]], N[ZetaZero[-x]]}, {x,1,20}]//TableForm` . Notice every value is of the form ".5+(some number) i"? There's the famous hypothesis in action. Every zero ever found has been of the form ".5+(some number) i". Given that the sum $\sum_\rho li(n^\rho)$ controls how much $\Pi(n)$ deviates from $li(n)$ , you can see how our knowledge about zeta zeros directly affects our knowledge about the long term behavior of $\Pi(n)$ .

A few things to note. One: this does indeed compute the Riemann prime counting function as advertised. Two: this is really, really complicated. And we're only scratching the surface here; to really understand the zeta zeros (in as much as people can) requires understanding the Riemann zeta function. And there's certainly no clues here as to why this formula works in the first place, either – to understand really does go deep down the rabbit hole of complex analysis. If you remain intrigued, though, I highly recommend Harold M. Edwards' "The Riemann Zeta Function".

Notice, too, with this explicit formula, if the Riemann Hypothesis is true,

$$\left|\sum_\rho li(n^\rho)\right| < \frac{1}{8\pi}\sqrt{n}\log n \quad \text{for } n \geq 59$$

Great stuff, but challenging. Let's move on to something more elementary.

---

## The First Elementary Approach: Smoothing

Take a look at this recursive function.

$$p_k(n,j,d)=\begin{cases}\frac{1}{d}\cdot\left(\frac{1}{k}-p_{k+1}\left(\frac{n}{j},1+\frac{1}{d},d\right)\right)+p_k\left(n,j+\frac{1}{d},d\right) & \text{if } n\geq j\\ 0 & \text{if } n<j\end{cases}$$

```
p[n_, j_, k_, d_] := If[ n<j, 0, (1/d)( 1/k - p[n/j, 1+1/d, k+1, d]) + p[n,j+1/d,k,d]]
```

Seems pretty simple, doesn't it – only a bit more complicated than, say, the very well-known recursive definitions for the factorial or the Fibonacci numbers.

And yet, with the right initial terms, it's the Riemann prime counting function.

$$p_1(n,2,1)=\Pi(n)$$

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[ n^(1/k)]/k, {k,1,Log[2,n]}]
p[n_,j_,k_,d_]:=If[n<j,0,(1/d)(1/k-p[n/j,1+1/d,k+1,d])+p[n,j+1/d,k,d]]
Table[{referenceRiemanPrimeCount[n], "=",p[n,2,1,1]},{n,2,100}]//TableForm
```

Actually, it's rather more than that. With slightly different initial terms, it's the logarithmic integral!

$$\lim_{d\to\infty} p_1\left(n, 1+\frac{1}{d}, d\right) = li(n) - \log\log n - \gamma$$

```
p=Compile[{{n,_Real},{j,_Real},{k, _Real},{d, _Real} },If[n<j,0,d(1./k-p[n/j,1+1/d,k+1,d])+p[n,j+1/d,k,d]],CompilationTarget->"C"];
z=.08;
Table[{n, N[LogIntegral[n]-Log[Log[n]]-EulerGamma], p[n,1+z,1,z]}, {n,2,7}]
```

Note that as *d* gets close to infinity the Mathematica code gets really, really, monstrously slow – errors in the approximation are because the value of d here isn't all that close to infinity. (Note that log log n grows at an utterly glacial rate. For example, $\log(\log(10^{100})) = 5.4392\ldots$ . The constant $\gamma = .577\ldots$ is negligible as well.)

So, the Riemann Hypothesis is equivalent to the following inequality being true.

$$\left| p_1(n, 2, 1) - \lim_{d\to\infty} p_1\left(n, 1+\frac{1}{d}, d\right) \right| < \frac{1}{8\pi}\sqrt{n}\log n \quad \text{for } n \geq 59$$

Viola!

Although the above definition is delightfully compact, it might be easier to get a sense of what's happening here if the recursion is manually flattened out. Doing that leads to

$$\Pi(n) = \sum_{j=2}^{n} 1 - \frac{1}{2}\sum_{j=2}^{n}\sum_{k=2}^{\lfloor \frac{n}{j} \rfloor} 1 + \frac{1}{3}\sum_{j=2}^{n}\sum_{k=2}^{\lfloor \frac{n}{j} \rfloor}\sum_{l=2}^{\lfloor \frac{n}{j\cdot k} \rfloor} 1 - \frac{1}{4}\sum_{j=2}^{n}\sum_{k=2}^{\lfloor \frac{n}{j} \rfloor}\sum_{l=2}^{\lfloor \frac{n}{j\cdot k} \rfloor}\sum_{m=2}^{\lfloor \frac{n}{j\cdot k\cdot l} \rfloor} 1 + \ldots$$

$$li(n) - \log\log n + \gamma = \int_1^n dx - \frac{1}{2}\int_1^n\int_1^{\frac{n}{x}} dy\, dx + \frac{1}{3}\int_1^n\int_1^{\frac{n}{x}}\int_1^{\frac{n}{xy}} dz\, dy\, dx - \frac{1}{4}\int_1^n\int_1^{\frac{n}{x}}\int_1^{\frac{n}{xy}}\int_1^{\frac{n}{xyz}} dw\, dz\, dy\, dx + \ldots$$

(For a general proof outline of this, see [http://icecreambreakfast.com/primecount/ApproximingThePrimeCountingFunctionWithLinniksIdentity_NathanMcKenzie.pdf](http://icecreambreakfast.com/primecount/ApproximingThePrimeCountingFunctionWithLinniksIdentity_NathanMcKenzie.pdf) .)

$$\zeta(s) = \lim_{n\to\infty} \sum_{j=1}^{n} j^{-s} - \int_0^n x^{-s}\, dx$$

$$\lim_{n\to\infty} \sum_{j=1}^{n} j^{-\rho} - \int_0^n x^{-\rho}\, dx = 0$$

So that's one way to connect the two functions.

---

## The Second Elementary Approach: Generalized Alternating Signs

Here's another elementary approach connecting $\Pi(n)$ to $li(n)$

If you're familiar at all with the Riemann zeta function, you almost certainly know its special relationship with the

Dirichlet eta function. The zeta function is defined as $\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}$ when the real part of s is greater than 1, and the eta function is defined as $\eta(s)=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^s}$ when the real part of s is greater than 0. The two functions are connected via the identities

$$\eta(s)=(1-2^{1-s})\zeta(s)$$

and

$$\zeta(s)=(1-2^{1-s})^{-1}\eta(s)$$

Consequently, we can say things about the eta function that in turn reveal properties of the zeta function.

Notice two details. First, one function has a numerator of 1 and the other a numerator that alternates positive/negative signs. Second, the number 2 seems to play a special role.

Anyway, we can use a similar approach here.

Like in the previous section, we start with a recursive function, but this time it leads with an alternating positive / negative sign:

$$f_k(n,j)=\begin{cases}(-1)^{j+1}\left(\frac{1}{k}-f_{k+1}\left(\frac{n}{j},2\right)\right)+f_k(n,j+1) & \text{if } n\geq j\\ 0 & \text{if } n<j\end{cases}$$

A little investigation shows this computes the Riemann counting function too, as:

$$\Pi(n)=f_1(n,2)+\sum_{j=1}^{\lfloor\frac{\log n}{\log 2}\rfloor}\frac{2^j}{j}$$

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[n^(1/k)]/k,{k,1,Log[2,n]}]
f[n_,j_,k_]:=If[n<j,0,(-1)^(j+1)(1/k-f[n/j,2,k+1])+f[n,j+1,k]]
Table[{n,referenceRiemanPrimeCount[n],"=",f[n,2,1]+Sum[ 2^j/j,{j,1,Floor[Log[2,n]]}]},{n,1,100}]//TableForm
```

This identity is similar to the one from previous section, only with alternating signs and that peculiar leading term $\sum_{j=1}^{\lfloor\frac{\log n}{\log 2}\rfloor}\frac{2^j}{j}$ which sticks out like a sore thumb, much as the 2 in $\zeta(s)=(1-2^{1-s})^{-1}\eta(s)$ sticks out.

So what is the connection between $(-1)^{j+1}$ and 2, do you suppose?

Well, to cut to the chase, take a look at this function:

$$t_{\frac{b}{d}}(n)=d\cdot\left(\lfloor\frac{n}{d}\rfloor-\lfloor\frac{n-1}{d}\rfloor\right)-b\cdot\left(\lfloor\frac{n}{b}\rfloor-\lfloor\frac{n-1}{b}\rfloor\right)$$

$$t_{\frac{b}{d}}(n)=\alpha_d(n)-\alpha_b(n)\text{ where }\alpha_b(n)=\begin{cases}b & \text{if } b|n\\ 0 & \text{if } b\nmid n\end{cases}$$

```
t[n_, b_, d_]:=d (Floor[n/d]-Floor[(n-1)/d])-b (Floor[n/b]-Floor[(n-1)/b])
Table[{n,t[n,2,1], t[n,3,2], t[n,3,5], t[n,7,5]},{n,1,30}]//TableForm
```

Run this Mathematica code and you'll notice t[n,2,1] cycles between 1 and -1. This isn't the only way to generalize

$(-1)^{j+1}$ , but it is an important one ( $\sum_{j=1}^{\infty} \frac{t_{\frac{b}{d}}(j)}{j} = \log \frac{b}{d}$ generalizes the well-known $\sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} = \log 2$ , for example). Let's use $t_c(n)$ to generalize our alternating series, with *c* being some rational constant such that $c = \frac{b}{d}, b > d$ . In particular, *d* is the denominator of our fraction. Then we have a new generalized function,

$$f_k(n,j) = \begin{cases} t_{\frac{b}{d}}(j \cdot d) \cdot \frac{1}{d} \cdot (\frac{1}{k} - f_{k+1}(\frac{n}{j}, 1 + \frac{1}{d})) + f_k(n, j + \frac{1}{d}) & \text{if } n \geq j \\ 0 & \text{if } n < j \end{cases}$$

with the corresponding identity

$$\Pi(n) = f_1(n, 1 + \frac{1}{d}) + \sum_{j=1}^{\lfloor \frac{\log n}{\log c} \rfloor} \frac{c^j}{j}$$

```
referenceRiemanPrimeCount[n_]:=Sum[PrimePi[n^(1/k)]/k,{k,1,Log[2,n]}]
t[n_, b_, d_]:=d (Floor[n/d]-Floor[(n-1)/d])-b (Floor[n/b]-Floor[(n-1)/b])
f[n_,j_,k_,b_,d_]:=If[n <j,0,(1/d) t[j d,b,d](1/k-f[n/j,1+1/d,k+1,b,d])+f[n,j+1/d,k,b,d]]
primecount[n_,c_]:=f[n,1+1/Denominator[c],1,Numerator[c],Denominator[c]]+Sum[c^j/j,{j,1,Floor[Log[n]/Log[c]]}]
Table[{n,referenceRiemanPrimeCount[n],primecount[n,3],primecount[n,5/2],primecount[n,7]},{n,5,100,5}]//TableForm
```

Alright. So now we have this identity based on a generalized notion of alternating series for some constant rational number *c*. Where exactly does that get us?

Now, let's supposes $c = 1 + \frac{1}{d}$ . What do you suppose happens as $\lim_{d \to \infty}$ ?

Well, it turns out our funny leading term, adjusted by just a bit, when that limit is taken, is the logarithmic integral!

$$\boxed{\lim_{c \to 1^+} -H_{\lfloor \frac{\log n}{\log c} \rfloor} + \sum_{j=1}^{\lfloor \frac{\log n}{\log c} \rfloor} \frac{c^j}{j} = li(n) - \log \log n - \gamma}$$

```
Table[{n,Chop[-HarmonicNumber[Floor[Log[n]/Log[c]]]+Sum[N[c^j/j],{j,1,Floor[Log[n]/Log[c]]}]/.c->1.00001],N[LogIntegral[n]-
Log[Log[n]]-EulerGamma]},{n,5,100,5}]//TableForm
```

Which means we've arrived at yet another elementary equation showing the precise and explicit difference between the Riemann prime counting function and the logarithmic integral.

$$\boxed{\Pi(n) = li(n) - \log \log n - \gamma + \lim_{d \to \infty} f_1(n,c) + H_{\lfloor \frac{\log n}{\log c} \rfloor}}$$

And that means that the Riemann Hypothesis is equivalent to the statement

$$\left| \lim_{d \to \infty} f_1(n,c) + H_{\lfloor \frac{\log n}{\log c} \rfloor} \right| < \frac{1}{8\pi} \sqrt{n} \log n \quad \text{for } n \geq 59$$

and, of course, vice versa.

If you find this at all interesting, many more ideas and details about this approach can be found in section 8 of http://www.icecreambreakfast.com/primecount/GeneralizedDivisorResults.pdf

…

about this ap

$$\zeta(s)=(1-2^{1-s})^{-1}\eta(s)$$

about this ap

$$\log\zeta(s)=\log\eta(s)-\log(1-2^{1-s})$$

about this ap

$$f_k(n,j)=\begin{cases}(-1)^{j+1}\cdot j^{-s}\cdot(\frac{1}{k}-f_{k+1}(\frac{n}{j},2))+f_k(n,j+1) & \text{if } n\geq j\\ 0 & \text{if } n<j\end{cases}$$

about this ap

$$\sum_{p^k\leq n}\frac{p^{-s}}{k}=f_1(n,2)+\sum_{j=1}^{\lfloor\frac{\log n}{\log 2}\rfloor}\frac{2^{j(1-s)}}{j}$$

about this ap

$$\log\zeta(s)=\log\eta(s)-\log(1-2^{1-s})$$

...

about this ap

$$\lim_{c\to 1^+}-H_{\lfloor\frac{\log n}{\log c}\rfloor}+\sum_{j=1}^{\lfloor\frac{\log n}{\log c}\rfloor}\frac{c^{j(1-s)}}{j}=li(n^{1-s})-\log\log n^{1-s}-\gamma$$

$$\lim_{n\to\infty}\sum_{j=1}^{\lfloor\frac{\log n}{\log c}\rfloor}\frac{c^{j(1-s)}}{j}=-\log(1-c^{1-s})$$

---

# Conclusion