Wiener-Ikehara Tauberian Theorem

John D Mangual

I found "analytic number theory" or "prime number theory" a frustrating subject. The discussions are are lengthy and disorganized and seem to lead nowhere.

Let's try to prove the Prime Number Theorem:

$$\frac{1}{x} \times \#\{p \leq x : p \text{ is prime}\} = \frac{1}{\log x}$$

The proof is like 40 pages and I lost track of everything. The first step, there is always a prime between $n < p < 2n$ by close examination of $\binom{2n}{n}$ and the Euclidean algorithm[1].

[1] See Zagier's **The First 50 Million Prime Numbers** `http://people.mpim-bonn.mpg.de/zagier/files/doi/10.1007/BF03039306/fulltext.pdf`

Then I found out about the Tauberian Theorems[2] which prime to make the proof faster and more motivated. What was Tauber's theorem anyway?

First, Abel's Theorem on when we can put $x = 1$ in series.

Titchmarsh's **Theory of Functions** says[3] we can let $x \to 1$ often:

$$\lim_{x\to 1} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_n$$

if the thing on RHS exists there is uniform convergence.

Textbook example is to show $\log 2 = 1 - \frac{1}{2} + \frac{1}{3} + \frac{1}{4} - \cdots = \sum \frac{(-1)^n}{n}$ proven by setting $x = 1$ in Taylor's formula or MacLaurin's:

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots$$

[2]

[3] "Theory of Functions" is a fancy word for "Complex Analyis" – also the book of Knopp.

Tauber's thorem goes the other way. If a series converges

$$\sum_{n=0}^{\infty} a_n x^n \to s$$

as $x \to 1$ we are concluding the same is true when $x = 1$.

$$\sum_{n=0}^{\infty} a_n \to s$$

To reiterate[4] we are comparing $x = 1$ with $x \to 1$ (or $x \to 1^-$).

The textbook says $\frac{1}{1+x} \to \frac{1}{1+1} = \frac{1}{2}$ but the Taylor expansion doesn't agree

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-x)^n \to 1 - 1 + 1 - 1 + \cdots =_? \frac{1}{2}$$

In order to make these statements true we have thrown around the symbol "=" to mean quite a few things[5].

[4]These are really loaded statements... it is as if series $\sum a_n x^n$ were a dime a dozen. Plugging $x = 1$ is going to be the obvious move and suddenly we are going to get the wrong conclusion. Doing it in reverse and generating examples off the topic of my head, from memory, is so much work.

[5]Using the Cesaro mean we could observe the partial sums are $1, 0, 1, 0, 1, 0, \ldots$ and split the difference at $\frac{1}{2}$.

The Wiener-Ikehara Tauberian theorem is the culmination of numerious re-arrangement tricks with divergent series:

> Les séries divergentes sont en général quelque chose de bien fatal et c'est une honte qu'on ose y fonder aucune démonstration.
>
> Divergent series are in general something fatal, and it is a disgrace to base any proof on them
>
> The divergent series are the invention of the devil, and it is a shame to base on them any demonstration whatsoever

GH Hardy wrote a book on divergent series where he evaluates the plausibility of using them in "rigorous calculations".

most calculations are quite sloppy and nobody really notices are cares[6]

[6]In the 20th century, divergent series return with a vengeneance, charging straight for you hard and fast. $1 + 1 + 1 + \cdots = -\frac{1}{2}$

statement

Suppose[7] that $a(u) : [0,\infty) \to \mathbb{R}$ is non-negative and increasing and that

$$\hat{a}(s) = \int_0^\infty e^{-us}\, da(u)$$

converges for all s with $\sigma > 1$ and that

$$\hat{a}(s) - \frac{c}{s-1}$$

extends to a continuous function on the closed half-plane $\sigma \geq 1$. Then as $x \to \infty$

$$\frac{1}{e^x}\int_0^x 1\, da(u) = c + o(1)$$

---

the measure $da(u)$ has a Laplace transform $\hat{a}(s)$ that diverges near $s = 1$. Now we compute the integeral of the constant function $\mathbf{1}$ and see how it behaves[8]

[7] Math language is so whimsical... we are letting this and supposing that... where does a function with such very specific properties come from? It is going to come from the prime numbers...

[8] I have re-written it slightly to make it look like an average. This is not looking like a statement on divergent series. And besides why be so rigorous since most people don't care anyway?

I should also mention the Karamata Theorem – which is also a Tauberian Theorem and leads to a proof of PNT – but requires more inputs and more footwork on our own.

Wiener-Ikehara Tauberian theorem will simply return the Prime Number Theorem as a corollary (among many others).

The statement the Tauberian Theorem is due to Montgomery and Vaughan. However, the Wiener in this article is Norbert Wiener the prodigy from MIT who helped formulate are modern language about stochastic processes.

Could it be that the primes are exhibiting white noise[9]?

---

[9] A rather innocuous observation about the density of primes took over a century to prove and

Suppose that $a_n \geq 0$ for all n and $s = \sigma + i\tau$ with $\sigma \geq 1$:

$$\hat{a}(s) = \sum_{n=1}^{\infty} a_n n^{-s} = \frac{c}{s-1} + \text{(something continuous)}$$

Then the average coefficient approaches a number:

$$\frac{1}{x} \sum_{n \leq x} a_n = c + o(1)$$

In our case we set $a_n = \Lambda(n)$ the **Van Mangoldt function** which is:

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \\ 0 & \text{otherwise} \end{cases}$$

The statement shows the average value $\Lambda(n)$ is just 1:

$$\frac{1}{x} \sum_{n \leq x} \Lambda(n) = 1 + o(1)$$

Strategy we wish to argue that $\Lambda(n) \approx 1$ despite that being wildly not true.

- for large primes p we have $\Lambda(n) = \log p \gg 1$
- $\Lambda(n)$ is often 0.
- These two facts average out perfectly to $\Lambda \approx 1$.

Another statement with the Möbius function $\mu : \mathbb{N} \to \{-1, 0, 1\}$.

$$\mu(n) = \left\{ \begin{array}{rl} 0 & \text{if } p^2 \text{ divides } n \\ 1 & \text{if } n = p_1 \ldots p_k \text{ with } k \text{ odd} \\ -1 & \text{if } n = p_1 \ldots p_k \text{ with } k \text{ even} \end{array} \right\}$$

Then we are trying to show $\mu(n) \approx 0$ even though 2/3 of the times it is ± 1.

$$\frac{1}{x} \sum_{n \leq x} \mu(n) = 0 + o(1)$$

Even though on average $\mu(n)$ is zero, there is obviously a lot of interesting behavior that happens quite often.

The main lemma involves the exponential funnction. Let $E(x)$ be e^x for negative numbers and 0 for positive numbers.

For any $\epsilon > 0$, there are continuous functions $f_{\pm}(x)$ approximating the exponential function:

- $f_-(x) \leq e^x \leq f_+(x)$ for all $x < 0$.
- $f_-(x) \leq 0 \leq f_+(x)$ for all $x > 0$.
- $\hat{f}_{\pm}(t) = 0$ for $|t| \geq T$ Fourier transform has compact support
- The integrals are approximately 1:

$$1 - \epsilon < \int_{-\infty}^{\infty} f_-(x)\,dx < 1 < \int_{-\infty}^{\infty} f_+(x)\,dx < 1 + \epsilon$$

Charts! I need lots and lots of charts! Let:

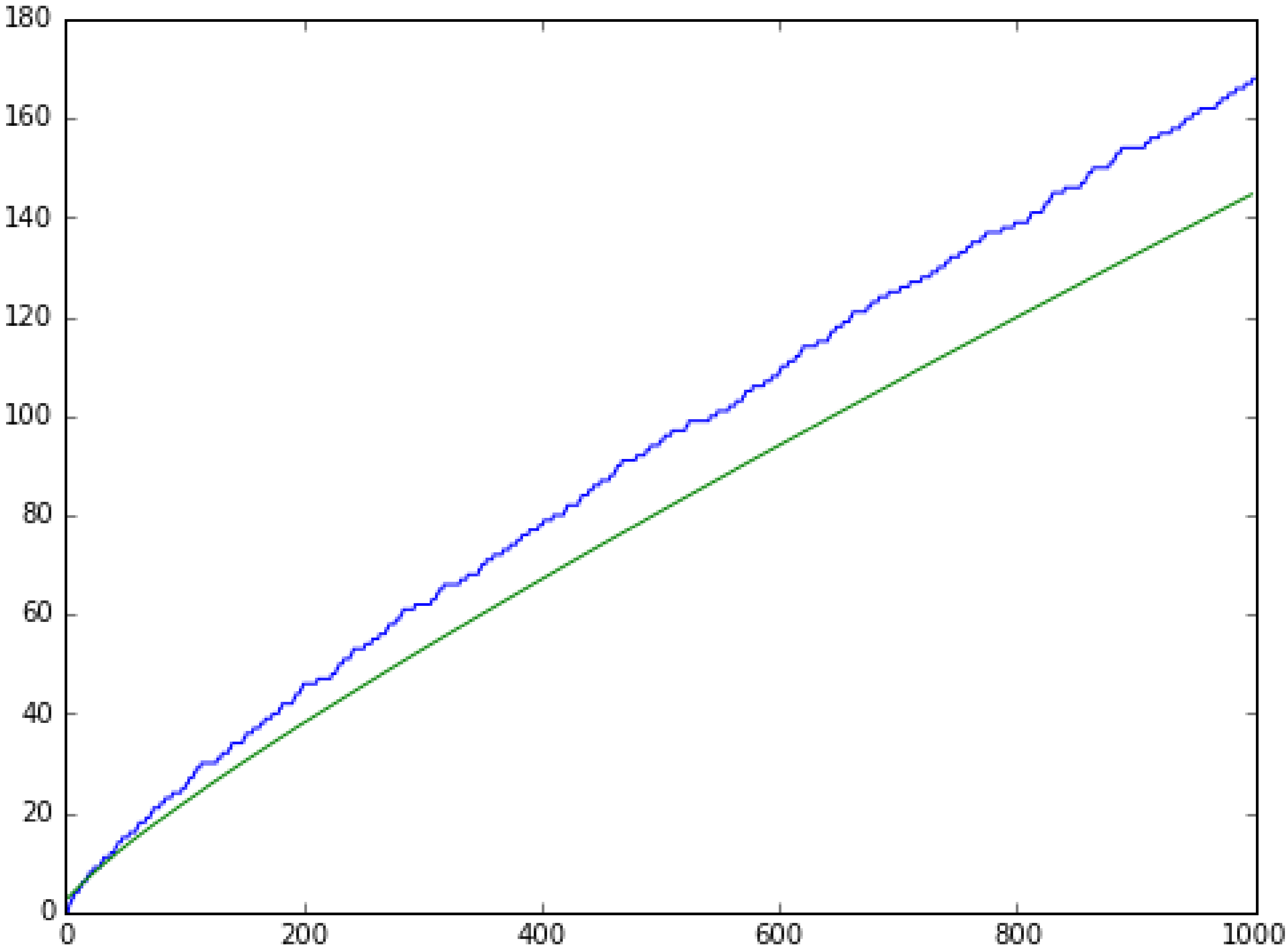
$$A(x) = \sum_{n \le x} \Lambda(n) \approx x + o(x)$$

The prime number theorem is the statement that $A(x) \approx x$ and the chat clearly shows as much.

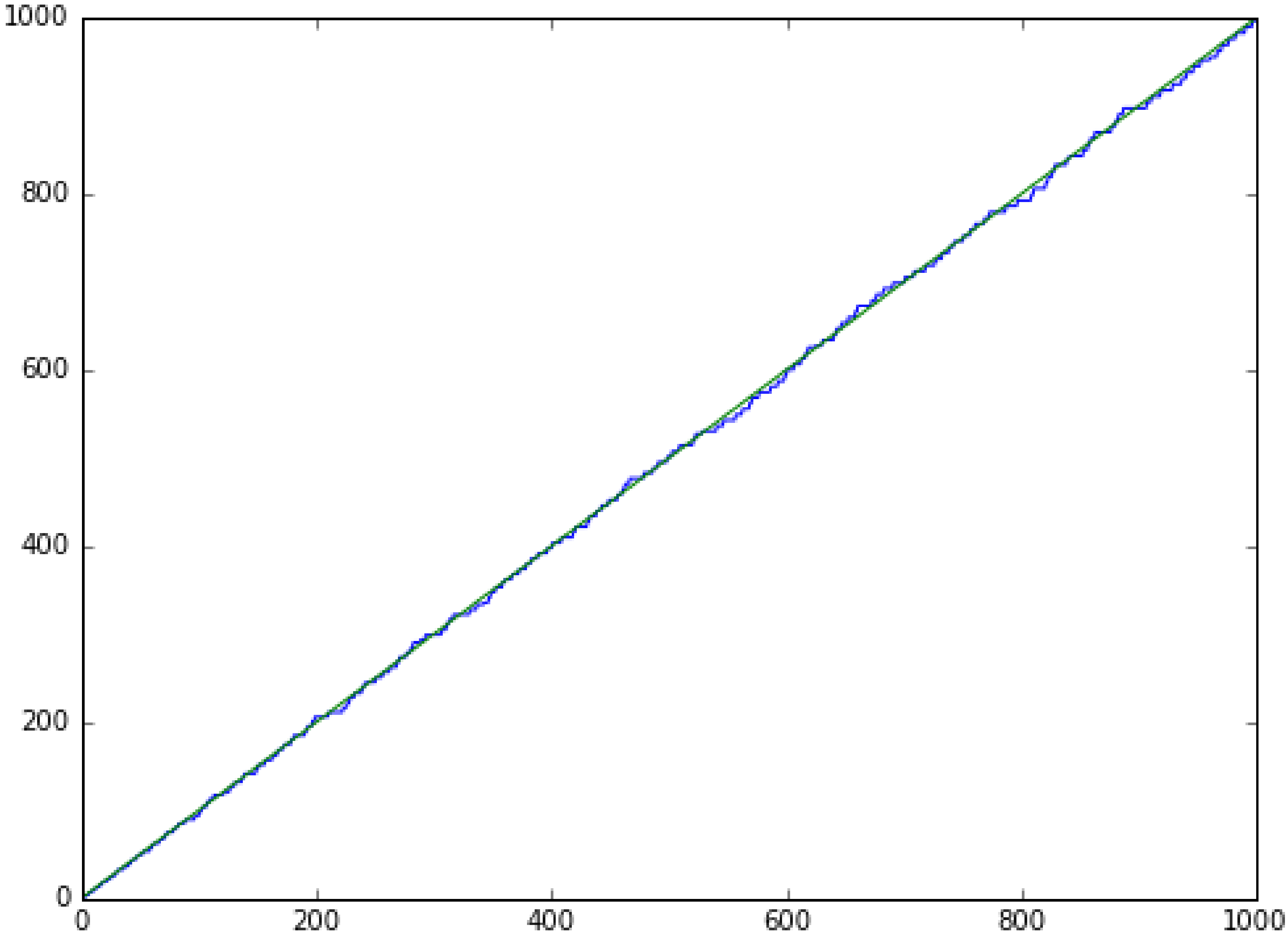
the **proof** requires showing the studying the error and showing it's small enough, and the truth is $\Lambda(n) - 1$ as a lot of interseting behavior (over all scales) so it's quite hard to show on average it is zero.

These primes! What are these primes like???

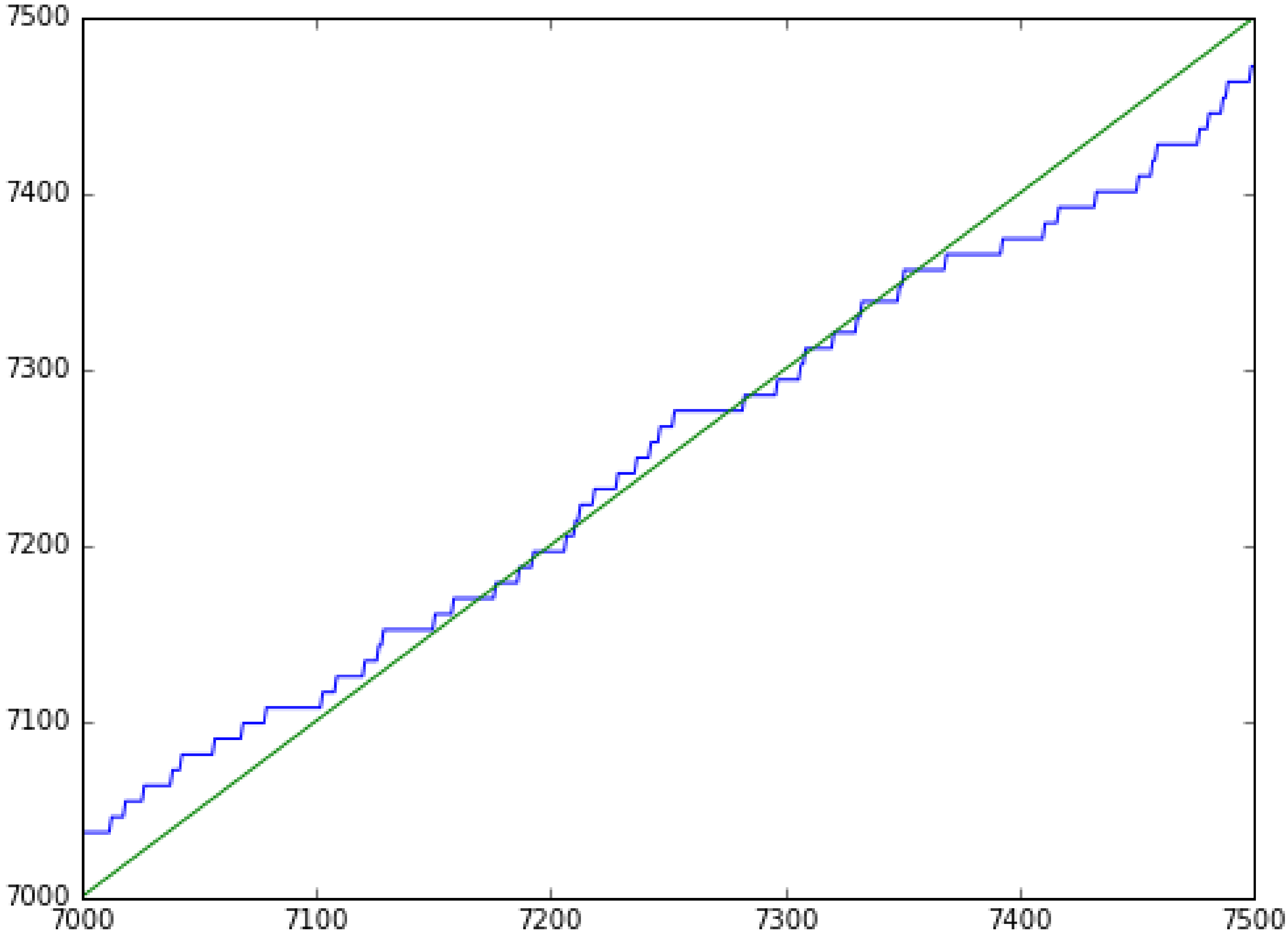
Compare $\#\{p < x\}$ vs $\frac{x}{\log x}$



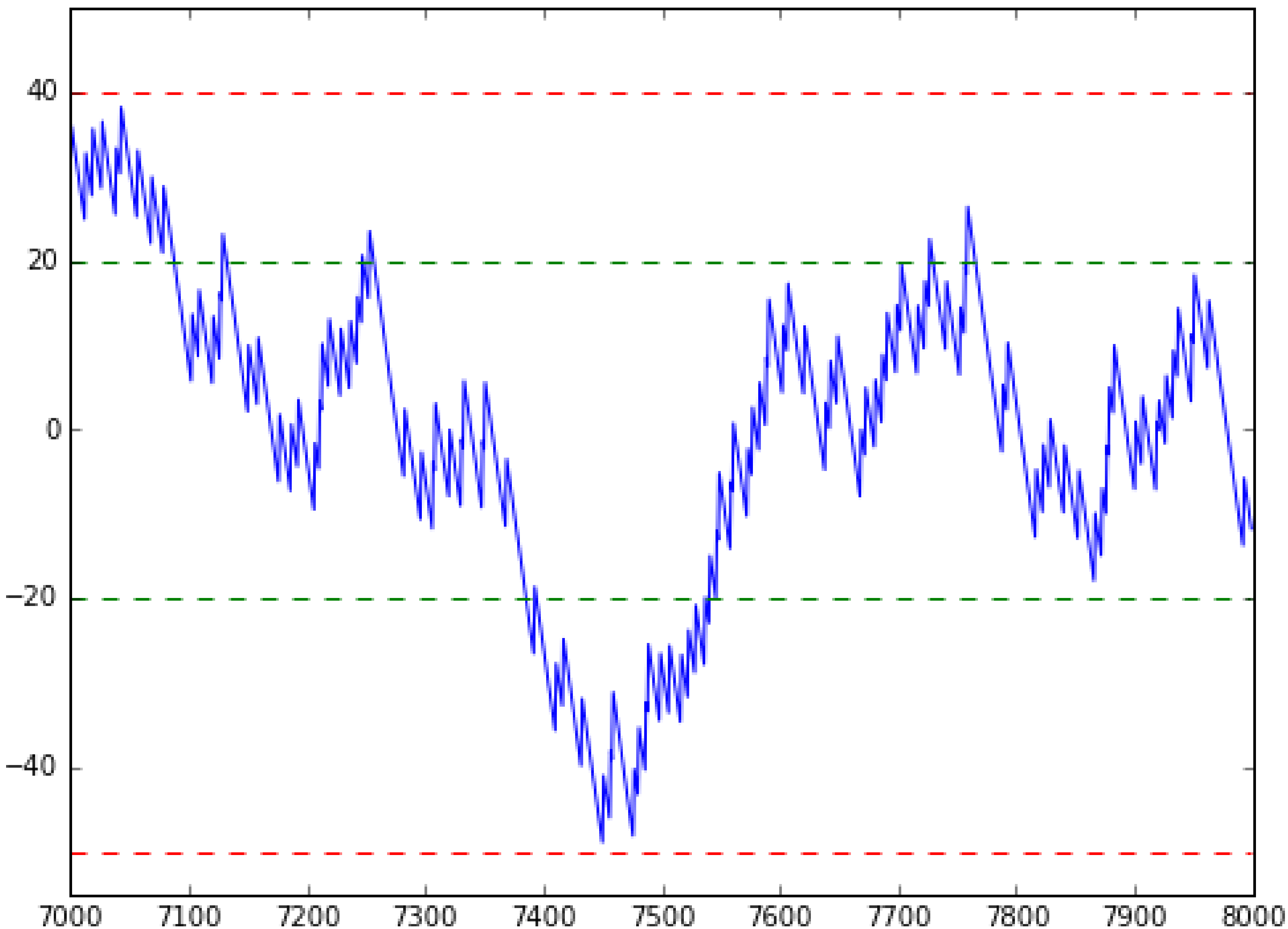
The perfect fit for $\sum_{n \le x} \Lambda(n) \approx x$



The perfect fit for $\sum_{n \leq x} \Lambda(n) \approx x$ (close up around $n = 7500$)



Approximating $\sum_{n\leq x}\Lambda(n) - x$ as random walk is not quite right[10]



[10]Neither deterministic nor totally random, more like a dynamical system converging to randomness (like shuffling a dominoes or a deck or cards).

In our case, the Laplace transform corresponds to something specific. Or a Mellin transform since we are doing $\times$ not $+$:

$$\sum_{n\le x}\Lambda(n)\,n^{-s}=\int_1^\infty v^{-s}d\left[\sum_{n\le x}\Lambda(n)\right]$$

The left hand side is $\frac{\zeta'(s)}{\zeta(s)}$ and the right side is a **Stieltjes integral**

Wiener-Ikehara theorem extends a single Laplace transform:

$$a(u)=e^u\to\hat{a}(s)=\int_0^\infty e^{-us}\,e^u\,du=\frac{1}{s-1}$$

The pole at $s=1$ "corresponds to" the distribution e^u on $\mathbb{R}^-$.

$$\int_0^\infty f_+(u-x)e^{-\delta u}du=\int_{-T}^{T}\hat{f}_+(t)e(-tx)\frac{1}{\delta-2\pi it}dt$$

We were told the mysterious f_+ kernel would only use specific wavelengths. I guess I'm OK with that.

Why does this remainder term tend to zero what could it mean?

$$e^x \int_{-T}^{T} \hat{f}_+(t)\, e(-tx) \left[\sum_{n=0}^{\infty} \Lambda(n)\, n^{-1+2\pi i t} - \frac{1}{2\pi i\, t} \right] dt \longrightarrow 0$$

This is the noisy part which, by the Riemann-Lebesgue lemma tends to zero.

Whay does the Riemann-Lebesgue Lemma[11] say? It says:

$$g \in L^1(\mathbb{R}) \longrightarrow \lim_{|\xi| \to \infty} \hat{g}(\xi) = 0$$

No wonder Montgomery-Vaughan's proof is so short!

Wikipeia says rigorous treatments of steepest descent and stationary phase are based on Riemann-Lebesgue Lemma. Essentially the integral over high enough frequencies should be zero since $\sin x$ is half the time $+1$ and half the time -1. This is the point of picking a function $f_(x)$ mimicking e^x, but whose Fourier coefficients $\hat{f}(t)$ stay within the range $-T < t < T$.

The magic function f is a far from obvious mix of the **Fejér kernel** and the lesser known Jackson Kernel. There might be others.

[11] We take from Katznelson **Harmonic Analysis**

References

(1) Hugh Montgomery, Robert Vaughan **Multiplicative Number Theory: 1. Classical Theorem** Cambridge, 2006.

(2) GH Hardy **Divergent Series**

(3) Titchmarsh **Theory of Functions** `https://archive.org/details/TheTheoryOfFunctions`

(4) Don Zagier **The First 50 Million Prime Numbers** New Mathematical Intelligencer 0 (1977) 1-19
`http://people.mpim-bonn.mpg.de/zagier/files/doi/10.1007/BF03039306/fulltext.pdf`