Introduction to Analytic Number Theory

via Dirichlet

A Reference Guide

Daniel K. Rui - September 23, 2022

Abstract

Recently (8/23/21 – see green 2021 math notebook) I became interested in getting a semi-serious first dose of analytic number theory (I do claim to have a deep enough interest in the subject to pursue graduate studies in it, but I have not actually seriously looked over it much detail).

1 Motivation

My starting point was 3b1b's video "Pi hiding in prime regularities", as it does introduce Dirichlet characters at the end and is an excellent problem for developing interesting ideas, and serving as a springboard for more complicated work. However, I did feel that the introduction of the Dirichlet character χ was just too slick/genius (as of now; let's hope someone writes a good answer to my MSE question), so I began looking into how Dirichlet came up with them. We proceed here introducing/motivating Dirichlet's work on prime in arithmetic progressions, mainly just leaving links and thoughts about those links (and I suppose the brilliant Gauss circle problem would be left as a supplementary/bonus section/exercise in this article).

Great article (https://arxiv.org/pdf/1404.4832.pdf – a lot of it is meta-mathematical/philosophy, but Section 3 talks about the math, and Appendix (pg. 48) talks about how Dirichlet may have come up with the idea of his characters, stemming from the study of Lagrange resolvents) detailing starting motivation of studying divergence of $\sum_{p\equiv a \bmod q}\frac{1}{p}$, by starting off with Euler product proof (we'll discuss that in a later section, Section 2) that there are infinitely many primes, i.e. proving that for $s>1$ (using Taylor expansion of $\log x$):

$$\log\left(\sum_{n=1}^{\infty}\frac{1}{n^s}\right)=\sum_{p}-\log\left(1-\frac{1}{p^s}\right)=\sum_{p}\frac{1}{p^s}+\sum_{n=2}^{\infty}\frac{1}{n}\sum_{p}\frac{1}{p^{ns}}=\sum_{p}\frac{1}{p^s}+\mathcal{O}(1)$$

where the constant $\mathcal{O}(1)$ is independent of s (which recall we are taking $s>1$) because

$$\left|\sum_{n=2}^{\infty}\frac{1}{n}\sum_{p}\frac{1}{p^{ns}}\right|\le\sum_{n=2}^{\infty}\frac{1}{n}\sum_{p}\frac{1}{p^{n}}\le\sum_{n=2}^{\infty}\frac{1}{n}\left(\int_{1}^{\infty}x^{-n}\,dx\right)=\sum_{n=2}^{\infty}\frac{1}{n}\cdot\frac{1}{n-1}\le\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty$$

Then breaking down

$$\log\left(\sum_{n=1}^{\infty}\frac{1}{n^s}\right)=\sum_{p}\frac{1}{p^s}+\mathcal{O}(1)=\sum_{a=1}^{q-1}\left(\sum_{p\equiv a \bmod q}\frac{1}{p^s}\right)+\mathcal{O}(1)$$

The hope is that we can write "similar formulas" to this line, where instead of summing over all a with weight 1, we have different weights so that via clever linear combinations we can isolate $\sum_{p\equiv a \bmod q} \frac{1}{p}$ on the RHS. These "similar formulas" would then hopefully be finite on the LHS, or at the very least not $-\infty$ at $s=1$ so that the divergence of $\log\left(\sum_{n=1}^{\infty} \frac{1}{n}\right)$ would lead to the divergence of $\sum_{p\equiv a \bmod q} \frac{1}{p}$ on the RHS. Writing this idea out, we want something like

$$\text{mysterious LHS, hopefully not } -\infty \text{ at } 1 = \sum_{a=0}^{q-1} w(a)\left(\sum_{p\equiv a \bmod q} \frac{1}{p^s}\right) + \mathcal{O}(1)$$

where w is a "weight function" defined on $\{0,\ldots,q-1\}$, mapping into anything $(\mathbb{Z},\mathbb{R},\mathbb{C})$ with the hope that with enough of these weight functions, doing linear combinations of them will result in cancellation in all but one $a\in\{0,\ldots,q-1\}$. In other words, we want enough $w:\{0,\ldots,q-1\}\to\mathbb{C}$ so that their span (in the vector space of these such functions) contain the functions $\delta_a:\{0,\ldots,q-1\}\to\mathbb{C}$ (1 at a, 0 everywhere else); note that as these δ_a form a basis of these such functions, this is equivalent to asking that our collection of w spans this vector space.

Also, note that in the $a=0$ case above, there is at most one prime $p\equiv 0 \bmod q \iff q$ divides p. Similarly, in the case that q is not prime and a,q share a factor other than 1 (i.e. $\gcd(a,q)\neq 1$), then there is again at most one prime $p\equiv a \bmod q$. Such a are not interesting to consider (the "at most one" prime can just be absorbed into the constant $\mathcal{O}(1)$, not impacting anything at all), so we can simply set $w(a)=0$. In other words, we really only care about the values of w on $a\in\{0,\ldots,q-1\}$ relatively prime to q. The set of such a is otherwise known as group of units $(\mathbb{Z}/q\mathbb{Z})^\times$ (though right now only thought of as a set). Sometimes we will abbreviate $(\mathbb{Z}/q\mathbb{Z})^\times$ as $\mathbb{Z}_q^\times$. We may then think of w as either a function $w:(\mathbb{Z}/q\mathbb{Z})^\times\to\mathbb{C}$, or as a q-periodic function $w:\mathbb{N}\to\mathbb{C}$ with zeroes at certain values; taking this latter viewpoint (i.e. the notion that w may be extended to all inputs $\mathbb{N}$), we may "bring the w inside" as follows:

$$\sum_{a\in\mathbb{Z}_q^\times}^{q-1} w(a)\left(\sum_{p\equiv a \bmod q} \frac{1}{p^s}\right) + \mathcal{O}(1) = \sum_{a\in\mathbb{Z}_q^\times}\left(\sum_{p\equiv a \bmod q} \frac{w(p)}{p^s}\right) + \mathcal{O}(1) = \sum_p \frac{w(p)}{p^s} + \mathcal{O}(1)$$

This looks very similar to the above sum $\sum_p \frac{1}{p^s} + \mathcal{O}(1) = \sum_p -\log\left(1-\frac{1}{p^s}\right)$, so is there a similar log expression for our weighted sum? Well, in that formula, we just used the Taylor expansion $-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots$ and took a sum over all primes p, meaning that if we set "x" to be instead $\frac{w(p)}{p^s}$, the sum over all primes $\sum_p \frac{w(p)}{p^s} + \mathcal{O}(1)$ (again higher order terms, even in $|\bullet|$, shrink fast enough like above, assuming w bounded, which is totally reasonable assumption — could even bound $|w|\leq 1$ by dividing by bound B) would simply be the sum over all primes $\sum_p -\log\left(1-\frac{w(p)}{p^s}\right)$.

Can we go one step further back, to something like $\log\left(\sum_{n=1}^{\infty}\frac{1}{n^s}\right)$? Recall that the Euler product formula was noting that $\prod_p (1-p^{-s})^{-1} = \prod_p(1+p^{-s}+(p^2)^{-s}+\ldots) = \sum_n n^{-s}$ because by the fundamental theorem of arithmetic every $n\in\mathbb{N}$ can be written as a product of primes in unique way $n = p_1^{e_1}\cdots p_r^{e_r}$, implying that every n on the RHS is attained exactly once (coefficient of n^{-s} for each $n\in\mathbb{N}$ equals 1). If we instead have $\prod_p(1-w(p)p^{-s})^{-1} = \prod_p(1+w(p)p^{-s}+w(p)^2(p^2)^{-s}+\ldots)$, again by the fundamental theorem of arithmetic each $n \in \mathbb{N}$ is attained on the RHS exactly once, except the coefficient of $n = p_1^{e_1} \cdots p_r^{e_r}$ is now $w(p_1)^{e_1} \cdots w(p_r)^{e_r}$.

In general (i.e. for general $w : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$), there is not much we can say about this product — basically this product takes the prime factorization of n and "hijacks" the primes p_i into $w(p_i)$, making the product highly dependent on the exact way n factors. However, this product LOOKS very similar to one that is much much simpler; SUPPOSING we could "pull" all these prime factors and exponents into one w, we would get very neatly $w(p_1)^{e_1} \cdots w(p_r)^{e_r} = w(p_1^{e_1} \cdots p_r^{e_r}) = w(n)$ (so going from an expression that requires knowledge of how exactly n factors, to an expression of just n). In this "dream scenario", we can write that for some (very hopefully extant!) "special" w,

$$\log\left(\sum_{n=1}^{\infty} \frac{w(n)}{n^s}\right) = \sum_p -\log\left(1 - \frac{w(p)}{p^s}\right) = \sum_p \frac{w(p)}{p^s} + \sum_{n=2}^{\infty} \frac{1}{n} \sum_p \frac{w(p^n)}{p^{ns}} = \sum_p \frac{w(p)}{p^s} + \mathcal{O}(1).$$

The property of "pulling everything in" that we really wanted w to have is simply multiplicativity, i.e. $w(mn) = w(m)w(n)$ for all $m, n \in \mathbb{Z}$. This allows us to pull the exponents e_i inside $w(p_i)^{e_i} = w(p_i^{e_i})$, and also allow us to pull the product over all prime factors inside. So, to summarize, we have found that multiplicative functions $w : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ seem to be a very promising direction of study in considering our question of primes in classes modulo q. It is remarkable that a condition as simple as multiplicativity allows us to write the extremely neat Euler product type formula above. MAJOR OBSERVATION: the multiplicative functions $w : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ are EXACTLY the group homomorphisms $(\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$, where we now think of $(\mathbb{Z}/q\mathbb{Z})^\times$ as a GROUP, not just a set. TIP: it's always a good sign in math when something from another field/problem shows up!

Using the fact that w is a group homomorphism $(\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$, we see that $w(1) = 1$ because for any $a \in (\mathbb{Z}/q\mathbb{Z})^\times$, $w(a) = w(a \cdot 1) = w(a)w(1)$. Lagrange's theorem can also be used to prove that $w(a)^{|\mathbb{Z}_q^\times|} =: w(a)^{\varphi(q)} = w(a^{\varphi(q)}) = w(1) = 1$ (as indicated by the "=:", I'm defining the Euler totient function $\varphi(q) := |\mathbb{Z}_q^\times|$). Thus, there is a cute notational coincidence because we just proved that $w(a)$ must be a root of unity in $\mathbb{C}$, which are usually denoted by some variant of the symbol "ω". However, in the literature, group homomorphisms $G \to \mathbb{C}$ are called *group characters*, and specifically the group homomorphisms $(\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ are called *Dirichlet characters*, and so from here on out, we will use χ instead of w. Finally, we note that the "mysterious LHS, hopefully not $-\infty$ at 1" that appeared in an above displayed equation is in fact $\log\left(\sum_{n=1}^{\infty} \frac{w(n)}{n^s}\right)$; defining the *Dirichlet L-function* $L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$, we see the "hopefully not $-\infty$ at 1" refers to the hope/conjecture that $L(1, \chi) \neq 0$ for Dirichlet characters χ.

From this much, we already see an OUTLINE. Somehow, we must understand these Dirichlet characters well enough to see if they EXIST first of all, and if they do, see if $\delta_a : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ can be written as a linear combination of such characters, or equivalently if these characters span (i.e. linearly combine into) the $\mathbb{Z}/\mathbb{R}/\mathbb{C}$-vector space of set-functions $(\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$. This is the CHARACTER THEORY portion of the proof. We then must understand the L-functions $L(s, \chi)$ well enough to determine that $L(1, \chi)$ is non-vanishing. This is the COMPLEX ANALYSIS portion.

2 Euler Product

Before we do all that, I want to say a few words on the Euler product. *On formulas that yield information about additive decomposition*. Easiest/earliest entry point is related to the identity $(a^2 - b^2) = (a-b)(a+b)$, more specifically $x^2-1 = (x-1)(x+1)$ and the geometric series. Upon seeing the formula $(a^2-b^2) = (a-b)(a+b) \implies x^2-1 = (x-1)(x+1)$, a high-schooler (such as myself; I do remember doing this) may see that with powers of 2 in exponent this formula can be iterated to yield $x^{2^n} - 1 = (x-1)(1+x)(1+x^2)(1+x^4)(1+x^8)\cdots(1+x^{2^{n-1}})$. One may recall from geometric series (literally the first series a high-schooler would learn) that $\frac{x^n-1}{x-1} = 1+x+x^2+\ldots+x^{n-1}$. Combining the two ideas/formulas/expressions we have

$$1+x+x^2+\ldots+x^{2^n-1} = \tfrac{x^{2^n}-1}{x-1} = (1+x)(1+x^2)(1+x^4)(1+x^8)\cdots(1+x^{2^{n-1}})$$

Already note of interest, we have a pure sum on the LHS and product on RHS. Also, every exponent $\{1,\ldots,2^n-1\}$ appears on the LHS with coefficient exactly 1. I think I can be confident in saying that even without understanding particularly deep math, this is very eye-catching at first glance. But why is this equation true (conceptual proof instead of algebraic manipulation)? When we multiply out the RHS, we see that this equation corresponds to the fact that every number $\{1,\ldots,2^n-1\}$ can be written in a unique way as 1/0-weighted sum of powers of 2 $\{1,2,\ldots,2^{n-1}\}$. In other words, I have n choices (i.e. digit places) where at choice i I choose between $+2^i$ or $+0$ (corresponding to n factors $(1+x)(1+x^2)(1+x^4)(1+x^8)\cdots(1+x^{2^{n-1}})$ where each factor has choice between x^{2^i} or x^0).

So exponents of this "dummy variable" x encode information about additive decomposition based on some "basis elements" $\{0 \text{ or } 1, 0 \text{ or } 2, 0 \text{ or } 4, \ldots\}$ (the "or" is exclusive, which we henceforth denote "xor" for "exclusive or"). The reason why I call x a "dummy variable" is because it is actually not doing anything/representing anything, it is just sort of serving as a "clothes line" on which the "clips" (the exponents of x) are interesting. Three more examples: for base-3 decomposition, we have "basis elements" $\{0 \text{ xor } 1 \text{ xor } 2, 0 \text{ xor } 3 \text{ xor } 2\cdot 3, 0 \text{ xor } 3^2 \text{ xor } 2\cdot 3^2, \ldots\}$, in other words $(1+x+x^2)(1+x^3+x^6)(1+x^9+x^{18})\cdots = 1+x+x^2+x^3+\ldots$ and so on (again coefficient one means unique decomposition in base-3).

Next two examples very similar: consider product $(1+x)(1+x^2)(1+x^3)\cdots$ corresponding to basis elements $\{0 \text{ xor } 1, 0 \text{ xor } 2, 0 \text{ xor } 3, \ldots\}$ results in a polynomial/formal power series (i.e. infinite clothes line) where coefficient of x^n is the number of ways to write n as sum of distinct numbers. For coefficient of x^n equal to number of ways to partition n into positive integers, we want for instance the option to have multiple 1's; however, if 1 appears in multiple "basis elements", say in three "basis elements", there are $\binom{3}{2}$ ways of choosing two 1's. So then all possible number of 1's must appear in one "basis element", and indeed as each partition can have either zero 1's XOR one 1's XOR two 1's XOR etc. we see that the basis $\{0 \text{ xor } 1 \text{ xor } 1{+}1 \text{ xor } \ldots, 0 \text{ xor } 2 \text{ xor } 2{+}2 \text{ xor } \ldots, \ldots\}$ corresponding to $(1+x+x^2+\ldots)(1+x^2+x^4+\ldots)(1+x^3+x^6+\ldots)\cdots$ and so on. In other words we have encoded the partition number $p(n)$ in the formula above, which we can write succinctly (think of it just as notation representing above product) as $\sum_{n=1}^{\infty} p(n)x^n = \prod_{n=1}^{\infty} \frac{1}{1-x^n}$.

Because multiplying x^m, x^n adds exponent, above techniques allow us to understand some combinatorics of additive decompositions in terms of coefficient of resulting formal power series. For multiplicative decomposition, exponent not useful, so we forget about the x entirely. But the idea/principle remains. We know that $n \in \mathbb{N}$ can be factored uniquely into product of primes, so in the spirit of our say base-2 additive decomposition above, we have for $\mathbb{N}$ a *multiplicative* decomposition based on some "basis elements" $\{1 \text{ xor } 2 \text{ xor } 2^2 \text{ xor } \ldots, 1 \text{ xor } 3 \text{ xor } 3^2 \text{ xor } \ldots, 1 \text{ xor } 5 \text{ xor } 5^2 \text{ xor } \ldots, \ldots\}$ corresponding to $(1+2+2^2+2^3+\ldots)(1+3+3^2+\ldots)(1+5+5^2+\ldots)(1+7+\ldots)\cdots = 1+2+3+4+\ldots$.

This makes sense from a conceptual point of view — for any $n \in \mathbb{N}$, going up on the LHS to a large enough but finite number of "basis elements" with large but finite "length" for each "basis element" (so like length l meaning $[1+2+2^2+\ldots+2^l]$), we see that n appears; and indeed no matter how many basis elements you take and how long you take each basis element n appears ONLY ONCE. At face value though both sides are ∞. No worries, we can still use this EXACT principle to encode this same fundamental theorem of arithmetic fact but with convergence on both sides; simply consider

$$\prod_p \frac{1}{1-\left[\frac{1}{p^2}\right]} = \prod_p \left(1 + \left[\frac{1}{p^2}\right] + \left[\frac{1}{p^2}\right]^2 + \ldots\right) = \sum_{n=1}^{\infty} \frac{1}{n^2}$$

Indeed, this works for any exponent n^{-s} for $s > 1$. There's probably more to say regarding convergence of sum and product, and rigorously proving this equality, but really this document is more of a guide than an actual account of the detailed proof.

Last remarks: if we truncate the above (two paragraphs up) series of prime power basis elements, like $\{1 \text{ xor } 2 \text{ xor } 2^2 \text{ xor } \ldots \text{ xor } 2^{e_1}, 1 \text{ xor } 3 \text{ xor } 3^2 \text{ xor } \ldots \text{ xor } 3^{e_2}, \ldots, 1 \text{ xor } p \text{ xor } p^2 \text{ xor } \ldots \text{ xor } p^{e_r}\}$, we get on the RHS the sum of all divisors of $2^{e_1} \cdot 3^{e_2} \cdots p^{e_r}$. In general, defining the sum-of-divisors function $\sigma(n) := \sum_{d|n} d$, we have $\sigma(p_1^{e_1} \cdots p_r^{e_r}) = \prod_{i=1}^r (1 + p_i + \ldots + p_i^{e_i}) = \prod_{i=1}^r \frac{p_i^{e_i+1}-1}{p_i-1}$.

3 Preliminary Concrete Examples

https://abel.math.harvard.edu/~elkies/M259.02/dirichlet.pdf has examples for $q = 4, 8$.

4 Character Theory (References)

Almost every source I looked at had a section on this. It's covered in first halves of https://math.mit.edu/classes/18.785/2015fa/LectureNotes17.pdf (18.785 Number theory I Lecture #17 Fall 2015 11/10/2015), https://fse.studenttheses.ub.rug.nl/17834/1/bMATH_2018_MintjesMW.pdf ("The Proof of Dirichlet's Theorem on Arithmetic Progressions and its Variations", M. W. Mintjes Bachelor's Project Mathematics, University of Groningen — beautifully typeset!). Basically a form of Fourier analysis, as the characters are orthogonal basis of $L^2(G)$ — see also https://www-users.cse.umn.edu/~garrett/m/mfms/notes_c/dirichlet.pdf ("Primes in arithmetic progressions" (April 12, 2011) Paul Garrett). For textbook, I'm sure Dummit & Foote has section on character theory; really any algebra/representation theory textbook.

5 Complex Analysis (References)

https://mathoverflow.net/a/26096/112504 provides sketch: assume $L(1, \chi) = 0$; then because $\zeta(s)$ has pole at $s = 1$, product $\zeta(s)L(s, \chi)$ should extend nicely past $s = 1$. Coefficients of Dirichlet series for this product are $c(n) = \sum_{d|n} \chi(d)$. Converge absolutely on $\mathrm{Re}(s) > 1$, but because $\zeta(s)L(s, \chi)$ extend analytically to $\mathrm{Re}(s) > 0$, the Dirichlet series converge conditionally for $\mathrm{Re}(s) > 0$ (LANDAU THEOREM). However, $c(n^2) > 0$, and $\sum_{n=1}^{\infty} c(n)n^{-1/2}$, leading to contradiction.

Requires knowledge that $\zeta(s), L(s, \chi)$ extend meromorphically/analytically to $\mathrm{Re}(s) > 0$. Refer to Terry Tao's Notes 2 of his 254A 2014 blog post; ctrl-F "meromorphic extension" and "holomorphic on this region". Basically the method is to prove the local uniform convergence of partial sums. Terry's whole Notes 2 is VERY long and chock full of information: https://terrytao.wordpress.com/2014/12/09/254a-notes-2-complex-analytic-multiplicative-number-theory/.

OVERVIEW/REVIEW OF MANY DIFFERENT PROOFS: https://fse.studenttheses.ub.rug.nl/17834/1/bMATH_2018_MintjesMW.pdf ("The Proof of Dirichlet's Theorem on Arithmetic Progressions and its Variations", M. W. Mintjes Bachelor's Project Mathematics, University of Groningen) referenced above in character theory section, has SECTION 4 devoted to proving the complex analysis portion in a different way then what author did in SECTION 3 (SECTION 4 proof similar to above sketched proof from MathOverflow). SECTION 5 reviews different proofs, pointing out strenghts/weaknesses.

More Terry: https://mathoverflow.net/a/29435/112504 and then blog post https://terrytao.wordpress.com/2009/09/24/the-prime-number-theorem-in-arithmetic-progressions-and-dueling-conspiracies/. More on analytic continuation of $\zeta(s)$: https://mathoverflow.net/questions/58004/how-does-one-motivate-the-analytic-continuation-of-the-riemann-zeta-function. For REALLY heavy duty machinery complex analysis, we have Terry's ENORMOUS blog post on functional equations + gamma, beta, digamma functions + reflection/duplication/multiplication + Poisson summation + Fourier inversion https://terrytao.wordpress.com/2014/12/15/254a-supplement-3-the-gamma-function-and-the-functional-equation-optional/.

https://math.uchicago.edu/~may/REU2012/REUPapers/LiAng.pdf ("DIRICHLET'S THEOREM ABOUT PRIMES IN ARITHMETIC PROGRESSIONS" Ang Li) contains proof of this complex analysis portion using a lot of really hands on bounding. May be nice approach if one doesn't want to introduce so much complex analysis theory.

Another approach of interest (though probably more complicated/messy) is Mathologer's "super sum" approach in his masterclass video "Numberphile v. Math: the truth about 1+2+3+...=-1/12".

6 Algebraic Number Theory

Part of the reason why I love the 3b1b video "Pi hiding in prime regularities" that I started off with in Section 1 is because all that work with factorization in $\mathbb{Z}[i]$ is the first step into algebraic number theory, and connections with quadratic fields, and Dirichlet's class number formula. I don't know anything about any of that now (maybe good reference is https://people.reed.edu/~jerry/361/lectures/iqclassno.pdf "THE DIRICHLET CLASS NUMBER FORMULA FOR IMAGINARY QUADRATIC FIELDS" Jerry Shurman Reed College), but I do know enough that I can see what a brilliant problem this Gauss circle problem is. All the more crucial that I find a good answer to my MSE question "How would one motivate/know to introduce the Dirichlet character in the formula for the number of lattice points on a circle of radius $\sqrt{N}$".

And about imaginary quadratic fields, I recently heard about a professor at CU, Katherine Stange (https://math.katestange.net/ — brilliant website, well designed, nice to look at, full of resources for students, visualizations, obviously the work of someone who loves teaching) who has some work on visualizing them: http://math.colorado.edu/~kstange/papers/Stange-short-exp.pdf ("Visualizing imaginary quadratic fields") with a tiny bit of explanation in the form of a Reddit thread https://www.reddit.com/r/math/comments/2xs4t7/visualising_complex_quadratic_number_fields/; see also her fully-fledged paper on the topic https://arxiv.org/pdf/1410.0417.pdf ("Visualising the arithmetic of imaginary quadratic fields").

7 Collected Links

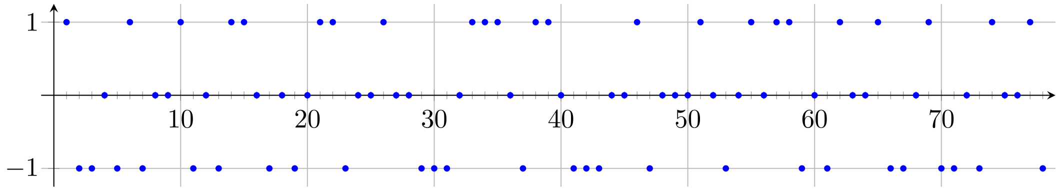
If one would like to have most of these above mentioned links in one list, see https://www.one-tab.com/page/uLC4YXuWTOWZxMWfER2_qg.

As concluding words, I mention Andrew Granville's excellent survey/overlook of basically the entire subject of analytic number theory (asymptotics, large/small prime gaps, sieve methods, circle method, Selberg/Langlands class of L-functions, etc.) https://dms.umontreal.ca/~andrew/PDF/PrinceComp.pdf.

8 The Möbius Function

It is 9/1/21 (everything prior written 8/25/21 - 8/26/21), and I have returned to add a bit on my MSE question "How would one motivate/know to introduce the Dirichlet character in the formula for the number of lattice points on a circle of radius $\sqrt{N}$". Although the answers/comments/pointers are very brief, we can spin up some ideas from them. First, given the Euler product formula $\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - \frac{1}{p^s})^{-1}$, we can easily (forgetting about convergence issues for now) invert and get an equally nice/concise formula $\frac{1}{\zeta(s)} = \prod_p (1 - \frac{1}{p^s})$.

Expanding out the RHS, we get $\frac{1}{\zeta(s)} = \prod_p (1 - \frac{1}{p^s}) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$ where $\mu(n)$ is 0 if the prime factorization of n has any exponent > 1, and otherwise if the prime factorization of n is $n = p_1^1 \cdot p_2^1 \cdots p_r^1$ then $\mu(n) = (-1)^r$. It's a bit crazy to me that the reciprocal of $\sum_{n=1}^{\infty} \frac{1}{n^s}$ is something that looks very very similar... like even though the formula for $\mu(n)$ is not super trivial (it is quite irregular, unlike the periodic Dirichlet characters discussed in previous sections), it's still amazingly simple for something that describes the multiplicative inverse of some infinite sum. This function $\mu(n)$ is the famous *Möbius function*. Let us plot its first couple values (taken from https://oeis.org/A008683/list):



Looking at this picture, the values of μ seem to be pretty evenly distributed between $-1, 0, 1$; indeed in the first 78 values of $\mu(n)$, the values $-1, 0, 1$ respectively appear $26, 29, 23$ times. A very natural question is to ask what proportion of n (as n tends to ∞) has $\mu(n) = -1, 0, 1$ respectively, i.e. consider the limits $\lim_{n\to\infty} \frac{|\mu^{-1}(-1,0,1) \cap [n]|}{n}$ (if they even exist). With our current data, one may reasonably suggest that maybe it's exactly even, that each of these limits is $\frac{1}{3}$.

A nice way of investigating this question is to compare the proportions of -1 and 1 by taking a running average $\frac{1}{N} \sum_{n=1}^{N} \mu(n)$. If the limit as $N \to \infty$ is positive, we know that there are slightly more 1's than (-1)'s, and if its negative, we know that there are slightly more (-1)'s than 1's. If it's 0, then we know the proportions are equal. Spoiler alert: it indeed turns out that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(n) = 0$ (i.e. the proportions between -1 and 1 are equal). This is already a very nice/aesthetically pleasing result, but the most shocking thing (at least to me, when I thought up this question today) is that this fact is EQUIVALENT to the prime number theorem!

With this result, we see that the partial sums $M(n) := \sum_{n=1}^{N} \mu(n)$ grow (in absolute value) at rate $o(N)$ (little-o of N). Determining the exact nature of how far $M(n)$ can deviate from 0 is a very complicated question — indeed the assertion that $M(n)$ is $\mathcal{O}(x^{1/2+\epsilon})$ for all $\epsilon > 0$ is EQUIVALENT to the Riemann hypothesis. Due to the disproof of Merten's conjecture (linked in the previous sentence) where in particular $\limsup_{n\to\infty} \frac{M(n)}{\sqrt{n}} > 1.8$ was shown in 2016 by Hurst, this hypothesized upper bound is extremely tight.

Anyways, knowing the proportions between -1 and 1 are equal, we still have yet to find the proportion of 0's. A good way to do this is to consider the running average $\frac{1}{N} \sum_{n=1}^{N} |\mu(n)|$. If our "exactly evenly $\frac{1}{3}$" conjecture is right, then this running average would approach $\to \frac{2}{3}$. I did some Googling and found a ResearchGate article with link http://dx.doi.org/10.5831/HMJ.2014.36.2.467 or in PDF form http://koreascience.or.kr/article/JAKO201418964310493.pdf that seems to claim that this limit is $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\mu(n)| = \frac{6}{\pi^2} \approx 0.608$, so slightly below our conjectured $\frac{2}{3}$. Actually Wolfram says that this limit appears in Landau 1974 pgs. 604-609.

Last remarks: like above with the Dirichlet characters where we could write

$$\sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} = \prod_p \left(1 + \frac{\chi(p)}{p^s} + \frac{\chi(p^2)}{p^{2s}} + \ldots\right),$$

we can write something similar $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \prod_p (1 + \frac{\mu(p)}{p^s} + \frac{\mu(p^2)}{p^{2s}} + \ldots) = \prod_p (1 - \frac{1}{p^s})$ because $\mu(p) = -1$ for any prime p and $\mu(p^e) = 0$ for any exponent $e \geq 2$. In other words, we could write this red equality because $\mu(p_1^{e_1} \cdots p_r^{e_r}) = \mu(p_1^{e_1}) \cdots \mu(p_r^{e_r})$; however we could not write something like the middle expression in the green equality above because that would require $\mu(p^e) = \mu(p)^e$, which is not true for the Möbius function (but is true for the Dirichlet characters from previous sections). A function f (like μ) that can decompose into its values at prime powers like $f(p_1^{e_1} \cdots p_r^{e_r}) = f(p_1^{e_1}) \cdots f(p_r^{e_r})$ is called *multiplicative*, and a function (like the Dirichlet characters) f that further decomposes into its values at pure primes like $f(p_1^{e_1} \cdots p_r^{e_r}) = f(p_1)^{e_1} \cdots f(p_r)^{e_r}$ is called *totally multiplicative*. The absolute value of the Möbius function (briefly discussed above) is also multiplicative, where $|\mu|(p) = 1$ and $|\mu|(p^e) = 0$ for $e \geq 2$.

9 Average Order

(EDIT 9/23/22: best read Sections 10 and later before this one. Section 11 can be regarded as more motivated/in-context version of this section). I got a bit sidetracked from my original goal of talking about the miraculous appearance of the Dirichlet character in the Gauss circle problem. I think the starting spark for this investigation (like the starting spark to Dirichlet's approach to prime in arithmetic progressions was the proof of the existence of infinitely many primes using the Euler product; or the starting spark of coming up with the Euler product was seeing how products of sums of terms could encode arithmetical/combinatorial information about those terms) is the question of what the average number of divisors of a number is (so again a running average taken over $n \in [N]$ for N large).

The number-of-divisors $\tau(n)$ (τ for German *Teiler*, divisors) is obviously defined as $\sum_{d|n} 1$. So, we are looking at the average number of divisors up to N, $\overline{\tau}_N := \frac{1}{N} \sum_{n=1}^{N} \tau(n) = \frac{1}{N} \sum_{n=1}^{N} \sum_{d|n} 1$. The brilliant insight (or "spark") is to notice that for each divisor d, approximately $\frac{N}{d}$ (or exactly $\lfloor \frac{N}{d} \rfloor$) $n \in [N]$ have that d as a divisor, meaning that $\sum_{n=1}^{N} \sum_{d|n} 1 = \sum_{d=1}^{N} \lfloor \frac{N}{d} \rfloor$. One can also think of this as a summation interchange, going from $\sum_{n=1}^{N} \sum_{d|n} 1$ to $\sum_{d=1}^{N} \sum_{d|n, n \in [N]} 1 = \sum_{d=1}^{N} \lfloor \frac{N}{d} \rfloor$.

We can bound this sum by $\sum_{d=1}^{N} (\frac{N}{d} - 1) \leq \sum_{d=1}^{N} \lfloor \frac{N}{d} \rfloor \leq \sum_{d=1}^{N} \frac{N}{d}$, or in other words defining the Nth harmonic number to be $H_N := 1 + \frac{1}{2} + \ldots + \frac{1}{N}$, we have $\frac{1}{N}(NH_N - H) \leq \overline{\tau}_N \leq \frac{1}{N}(NH_N)$. We thus get that $\lim_{N \to \infty} \frac{\overline{\tau}_N}{H_N} = 1$. Another notation for this is $\overline{\tau}_N \sim H_N$. Interpreting H_N as a Riemann sum for the integral $\int_1^{N+1} \frac{1}{x}\, dx = \log(N+1)$, the slivers of area $1 \cdot \frac{1}{n} - \int_n^{n+1} \frac{1}{x}\, dx$ can be translated horizontally until they all lie in $[0,1]$ without overlap, meaning $H_N - \log(N+1)$ is an increasing sequence of real numbers bounded above by 1, meaning the limit $\gamma := \lim_{N \to \infty} H_N - \log(N+1) = \lim_{N \to \infty} H_n - \log(N) + \log(\frac{N}{N+1}) = \lim_{N \to \infty} H_n - \log(N)$ exists (the Euler-Mascheronic constant), and so obviously $\lim_{N \to \infty} \frac{H_N}{\log N} = 1 \iff H_N \sim \log N$. Putting everything together, we get that $\overline{\tau}_N \sim \log N$, or the average order of $\tau(n)$ is $\log n$.

The reason this is so profound (or the “spark” of an entire investigation) is because the “1” in the above sum $\sum_{d|n} 1$ was not special at all; in fact we can generalize this to any function f: if $g(n) = \sum_{d|n} f(d)$, then $\overline{g}_N := \frac{1}{N}\sum_{n=1}^N \sum_{d|n} f(d) = \frac{1}{N}\sum_{d=1}^N f(d)\lfloor \frac{N}{d} \rfloor \approx \frac{1}{N}\sum_{d=1}^N f(d)\frac{N}{d} = \sum_{d=1}^N \frac{f(d)}{d}$. We can give more quantitative information about how exactly this is approximated (i.e. the symbol “$\approx$” is quite vague) as follows:

$$\left|\sum_{d=1}^N \frac{f(d)}{d} - \overline{g}_N\right| \le \frac{1}{N}\sum_{d=1}^N \left|f(d)\left(\frac{N}{d} - \left\lfloor \frac{N}{d} \right\rfloor\right)\right| = \frac{1}{N}\sum_{d=1}^N |f(d)| \cdot \left|\frac{N}{d} - \left\lfloor \frac{N}{d} \right\rfloor\right| \le \frac{1}{N}\sum_{d=1}^N |f(d)| \cdot 1 = \overline{|f|}_N.$$

If for instance f is bounded (so $|f| \le B$ for all inputs n), then the RHS would be bounded by B. Such a constant would be negligible if $\overline{g}_N \to \infty$ as $N \to \infty$, like in our above example where $\overline{\tau}_N \sim \log N \to \infty$ as $N \to \infty$.

This is where we return to my MSE question. In the notation in that question, we have that the total number of lattice points at distance $\le \sqrt{N}$ from the origin (forgetting about the origin itself, as 3b1b does in his video) is $\sum_{n=1}^N \#(n)$. Having primed our brains with sums of the form $\frac{1}{N}\sum_{n=1}^N g(n)$ above, we see that if we could express $\#(n)$ as $\sum_{d|n} f(n)$ for some other function f, we can apply our above knowledge to get something promising for the formula for the total number of lattice points within a circle of radius $\sqrt{N}$.

10 Möbius Inversion

Today is 2/7/22, and I suddenly remembered this after glancing over 246B Notes 4, which we are going to cover soon. I think I have given decent motivation for the Riemann zeta function and the Euler product formula in Section 2, so that assuages my concerns about 246B Notes 4 not sufficiently motivating the right-off-the-bat definition of $\zeta(s)$. I do remember wanting to say more about Möbius inversion back in September 2021, but I guess I never got around to it. Having introduced the Möbius function in a natural way from the Euler product formula, I think it is more natural to build up the inversion formula, before doing the above Section 9, since sums over divisors naturally show up in the inversion formula.

First, recall that $\frac{1}{\zeta(s)} = \prod_p (1 - \frac{1}{p^s}) = \sum_{n=1}^\infty \frac{\mu(n)}{n^s}$ and $\zeta(s) = \prod_p (1 - \frac{1}{p^s})^{-1} = \sum_{n=1}^\infty \frac{1}{n^s}$. Of course from the product representations it’s clear that these multiply to 1, but it is not at all obvious from the sum representation. But now that we know it’s true, we simply have to multiply out the series term by term, and see how things work out (should result in identically 1!). Because $\frac{a}{n^s} \cdot \frac{b}{m^s} = \frac{ab}{(nm)^s}$, the product will remain a series of the same form as $\zeta(s)$, $\frac{1}{\zeta(s)}$; such series are given the name *Dirichlet series*. The way to produce a $\frac{\bullet}{n^s}$ term in the product is to multiply $\frac{a}{d^s}$ and $\frac{b}{(n/d)^s}$ for any divisor $d \mid n$. In other words, the coefficient of $\frac{1}{n^s}$ in the product is $\sum_{d|n} \mu(d) \cdot 1$, or more generally the $\frac{1}{n^s}$ coefficient of the product of two Dirichlet series $\sum_{n=1}^\infty \frac{f(n)}{n^s}, \sum_{n=1}^\infty \frac{g(n)}{n^s}$ is $\sum_{d|n} f(d)g(\frac{n}{d})$. Commutativity suggests that $\sum_{d|n} f(d)g(\frac{n}{d}) = \sum_{d|n} g(d)f(\frac{n}{d})$, and indeed one can see this directly from $\sum_{d|n}$ by symmetry traversing the sum “forwards” i.e. d starts at 1 and goes up to n, vs. “backwards” i.e.

d starts at n and goes down to 1.

Also note that this operation of taking f, g defined on $\mathbb{N}$ to $\sum_{d|n} f(d)g(\frac{n}{d})$ (again defined for every $n \in \mathbb{N}$) looks a lot like convolution (a convolution integral is $\int f(x)g(y-x)\,dx$, and discretized it might look like $\sum_{k=0}^{n} f(k)g(n-k)$, which is basically our formula except with a distinct multiplicative flavor from $n-k \rightsquigarrow \frac{n}{d}$ and $\sum_{k=0}^{n} \rightsquigarrow \sum_{d|n}$). This is pretty good justification for using the convolution notation $[f*g](n) := \sum_{d|n} f(d)g(\frac{n}{d})$. So the above Dirichlet series product can be written succinctly as

$$\left(\sum_{n=1}^{\infty} \frac{f(n)}{n^s}\right) \cdot \left(\sum_{n=1}^{\infty} \frac{g(n)}{n^s}\right) = \sum_{n=1}^{\infty} \frac{[f*g](n)}{n^s}.$$

We also pointed out above that this "product" operation is commutative, and similarly it inherits associativity from just multiplication of numbers (although we are multiplying infinite sums, for any fixed $\frac{1}{n^s}$ term, only a finite number of multiplications/additions go into that coefficient). Deserving of the title "product" because it does distribute over addition: $[(f+g)*h](n) = \sum_{d|n}[f+g](d)h(\frac{n}{d}) = \sum_{d|n}\big(f(d)h(\frac{n}{d}) + g(d)h(\frac{n}{d})\big) = \sum_{d|n} f(d)h(\frac{n}{d}) + \sum_{d|n} g(d)h(\frac{n}{d}) = [f*h](n) + [g*h](n)$.

Anyways, because $(\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}) \cdot (\sum_{n=1}^{\infty} \frac{1}{n^s}) \equiv 1$ (for any $s > 1$), we have that $\sum_{d|n} \mu(d) = 0$ for all $n > 1$, and $= 1$ for $n = 1$, so denoting $\mathbf{1}$ to be the function identically 1 on $\mathbb{N}$, and δ to be the spike function 1 at $n = 1$ and 0 elsewhere, this can be written $\mu * \mathbf{1} = \delta$. These functions $\mathbf{1}$ and δ look particularly simple (especially the latter one), so let us see how they "convolve" with some arbitrary f: $[f * \mathbf{1}](n) = \sum_{d|n} f(d)$, and $[f*\delta](n) = \sum_{d|n} f(d)\delta(\frac{n}{d}) = f(n)\delta(1) + 0 + \ldots + 0 = f(n)$. To summarize, convolving with $\mathbf{1}$ is "summing over divisors", and δ is the identity for the "$*$" product operation. Therefore, combining this with associativity and commutativity, we get for any f: $f = f * \delta = f * (\mathbf{1} * \mu) = \mu * (f * \mathbf{1})$, so if $g = f * \mathbf{1} = \sum_{d|n} f(d)$, then one can recover f from g via μ: $f = \mu * g = \sum_{d|n} \mu(d) g(\frac{n}{d})$! This is the *Möbius inversion formula.*

11 Summing Over Divisors

More written 2/12/22. The Möbius inversion formula tells us there is a very natural way to represent any arithmetic function as the sum over divisors of some other arithmetic function (divisor sums/Möbius inversion overlays space of arithmetic functions on top of itself). Thus, a promising direction in which to travel would be to try to understand the behaviors of these sums $\sum_{d|n}$. Unfortunately, divisors behave in an extremely irregular way — looking at the rows of the below table, there seems to be no discernible pattern going from one value of n to the next.

$F(n) \setminus d$	1	2	3	4	5	6	7	8	9	10	11	12
$F(1) =$	$f(1)$											
$F(2) =$	$f(1)$	$f(2)$										
$F(3) =$	$f(1)$		$f(3)$									
$F(4) =$	$f(1)$	$f(2)$		$f(4)$								
$F(5) =$	$f(1)$				$f(5)$							
$F(6) =$	$f(1)$	$f(2)$	$f(3)$			$f(6)$						
$F(7) =$	$f(1)$						$f(7)$					
$F(8) =$	$f(1)$	$f(2)$		$f(4)$				$f(8)$				
$F(9) =$	$f(1)$		$f(3)$						$f(9)$			
$F(10) =$	$f(1)$	$f(2)$			$f(5)$					$f(10)$		
$F(11) =$	$f(1)$										$f(11)$	
$F(12) =$	$f(1)$	$f(2)$	$f(3)$	$f(4)$		$f(6)$						$f(12)$

However, the *columns* of the above table behave *extremely* regularly, in that every n is divisible by 1, every 2nd n is divisible by 2, every 3rd n is divisible by 3, and so on. Thus, although for just single values of n it is difficult to understand the behavior of the divisors, over multiple values of n the regularity of the rows might be able to help. In other words, the individualized behavior of divisors for any given n may be hard to understand, but the *average* behavior over the divisors of n over all $n \in [1, N] = \{1, \ldots, N\}$ is approximately that 1 will contribute all the time, 2 will contribute about half the time, 3 will contribute about a third of the time, and so on. More rigorously, the previous sentence says that the average $\frac{1}{N}\sum_{n=1}^{N} F(n)$, although difficult to analyze when summed over the rows first and then the columns, becomes much easier when summed over the columns first and then the rows, yielding

$$\begin{aligned}\overline{F([N])} := \frac{1}{N}\sum_{n=1}^{N} F(n) = \frac{1}{N}\sum_{n=1}^{N}\sum_{d|n} f(d) = \frac{1}{N}\sum_{d=1}^{N}\sum_{n\in[N]:d|n} f(d) = \frac{1}{N}\sum_{d=1}^{N} f(d)\left\lfloor\frac{N}{d}\right\rfloor \\ \approx \frac{1}{N}\sum_{d=1}^{N} f(d)\frac{N}{d} = \sum_{d=1}^{N}\frac{f(d)}{d}.\end{aligned}$$

We can give more quantitative information about how exactly this is approximated (i.e. the symbol "≈" is quite vague) as follows:

$$\begin{aligned}\left|\sum_{d=1}^{N}\frac{f(d)}{d} - \overline{F([N])}\right| &\le \frac{1}{N}\sum_{d=1}^{N}\left|f(d)\left(\frac{N}{d} - \left\lfloor\frac{N}{d}\right\rfloor\right)\right| = \frac{1}{N}\sum_{d=1}^{N}|f(d)|\cdot\left|\frac{N}{d} - \left\lfloor\frac{N}{d}\right\rfloor\right| \\ &\le \frac{1}{N}\sum_{d=1}^{N}|f(d)|\cdot 1 = \overline{|f|([N])}.\end{aligned}$$

Therefore, if $\overline{|f|([N])} = \mathfrak{o}\big(\overline{F([N])}\big)$ or $= \mathfrak{o}(\sum_{d=1}^{N}\frac{f(d)}{d})$, then $\sum_{d=1}^{N}\frac{f(d)}{d} \sim \overline{F([N])}$ as $N \to \infty$ (or equivalently $N\sum_{d=1}^{N}\frac{f(d)}{d} \sim \sum_{n=1}^{N} F(n)$, since $a(N) \sim b(N) \iff \lim_{N\to\infty}\frac{a(N)}{b(N)} = 1 \iff \lim_{n\to\infty}\frac{Na(N)}{Nb(N)} = 1 \iff Na(N) \sim Nb(N)$).

For instance, taking $f = \mathbf{1}$, the average value of f, i.e. 1 is indeed little-o of $\sum_{d=1}^{N} \frac{1}{d} \sim \log N$, so the above formula gives that the average number of divisors $d(n) := \sum_{d|n} 1$ of an integer in $[N]$ grows as $\log N$. To emphasize how "nicely" $d(n)$ behaves on average compared to for individual values of n, observe that $\liminf_{n\to\infty} d(n) = 2$ (because there are infinitely many prime numbers, and the only divisors of a prime number are 1 and itself), and $\limsup_{n\to\infty} d(n) = \infty$ (given the prime factorization of $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, the number of divisors is exactly $d(n) = (\alpha_1 + 1) \cdots (\alpha_k + 1)$, since there are $(\alpha_i + 1)$ choices for the exponent of each p_i, so taking $n = 2^{\alpha}$ for huge $\alpha \rightsquigarrow d(n) = \alpha + 1$ shows that the maximum number of divisors grows at least $\geq \log_2(n)$).

This already shows that $d(n)$ can vary quite wildly generally all over $\mathbb{N}$, but more specifically, we can show that $d(n)$ varies quite wildly just from one n to the next! Let Π_K denote the product of the first K primes $\rightsquigarrow d(\Pi_K) = 2^K$. Dirichlet's theorem on primes in arithmetic progressions tells us that there are infinitely many primes of the form $\Pi_K + 1$, so in fact for any K there are infinitely many n s.t. $|d(n+1) - d(n)| = 2^K - 2$. In particular, $\limsup_{n\to\infty} |d(n+1) - d(n)| = \infty$. In 1984, Roger Heath-Brown showed that $\liminf_{n\to\infty} |d(n+1) - d(n)| = 0$.

This section motivates the idea of studying averages of functions defined as sums over divisors, and introduces a technique (the Fubini trick, or interchanging rows and columns in a double sum) that can be used to understand such averages for a decent number of arithmetic functions. Because of the usefulness of this technique, a good strategy to study sums $\sum_{n\leq N} F(n)$ would be to use Möbius inversion to find f s.t. $F(n) = \sum_{d|n} f(n)$, and then use the Fubini trick. In particular, this section presents the best motivation I can come up with (at this point) for the material of Section 9 (especially that bit at the end where I talk about the 3b1b Gauss circle problem).

12 Upper Bound for Divisors

So we've talked about the the average behavior of $d(n)$, and the best-case behavior of $d(n)$ (2 for $n = p$ prime), but we have not talked about the worst-case behavior (only mentioned that the worst-case of $d(n)$ grows at least $\log_2(n)$, but no mention of upper bound). Obviously, we have $d(n) \leq n$. We can do a bit better by noticing that divisors come in pairs $(d, \frac{n}{d})$ where the first smaller number is always $\leq \sqrt{n}$, meaning the number of such pairs is $\leq \sqrt{n} \rightsquigarrow d(n) \leq 2\sqrt{n}$ (this is the same reasoning behind the easiest optimization of the most naïve primality checking algorithm, going from checking all $k \leq n$ to see if they divide n, to checking all $k \leq \sqrt{n}$).

Let us try to prove this another way. I already pointed out that for $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, the number of divisors is exactly $d(n) = (\alpha_1 + 1) \cdots (\alpha_k + 1)$, so our upper bound can be rephrased as

$$\prod_{i=1}^{k} \frac{\alpha_i + 1}{p_i^{\frac{1}{2}\alpha_i}} \leq 2.$$

Intuitively, this makes sense because the numerator grows linearly in α_i, while the denominator grows exponentially. Expanding on this idea to give a proof for the $n^{1/2}$ case essentially leads straight to the n^ϵ case; see https://terrytao.wordpress.com/2008/09/23/the-divisor-bound/.