

# Random number theory

Trjitzinsky Memorial Lectures II,  
University of Illinois Urbana–Champaign,  
November 28, 2018

Carl Pomerance

Dartmouth College (emeritus)

University of Georgia (emeritus)

In 1770, **Euler** wrote:

*“Mathematicians have tried in vain to discover some order in the sequence of prime numbers, but we have every reason to believe that there are some mysteries which the human mind will never penetrate.”*

from A. Granville, “Harald Cramér and the distribution of prime numbers”

In 1770, **Euler** wrote:

*“Mathematicians have tried in vain to discover some order in the sequence of prime numbers, but we have every reason to believe that there are some mysteries which the human mind will never penetrate.”*

Nevertheless, **Euler** proved in 1737 that the sum of the reciprocals of the primes to  $x$  diverges to infinity like  $\log \log x$ . So, 33 years before his pessimistic statement, he had a glimmer that the mysterious primes might obey some statistical law.

Less than 30 years after **Euler** opined on the mysteries of the primes, **Gauss**, as a teenager, arrived at the conjecture that the number of primes up to  $x$  is approximately

$$\int_2^x \frac{dt}{\log t}.$$

He wrote in 1849 in a letter to Encke:

*“As a boy I considered the problem of how many primes there are up to a given point. From my computations, I determined that the density of primes near  $x$  is about  $1/\log x$ .”*

op. cit.

Here are some notes in Gauss's hand found in the Göttingen library.

Yuri Tschinkel, courtesy of Brian Conrey

Math 18

② 44 of 2

# Primzahlen

von 100000 bis 110000.

	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	
1.	1.									1.	
2.		1.				1.		1.	1.	4.	
3.		4.	2.	2.	3.	1.	2.	3.	3.	21.	
4.	2.	8.	5.	4.	3.	6.	9.	4.	5.	54.	
5.	11.	10.	8.	18.	12.	10.	10.	12.	15.	8 114	
6.	14.	14.	18.	21.	16.	22.	19.	15.	17.	15. 171.	
7.	26.	17.	23.	23.	24.	24.	17.	22.	20.	21. 217.	
8.	19.	19.	21.	7.	14.	15.	20.	17.	15.	17. 164.	
9.	11.	13.	9.	13.	14.	14.	12.	13.	11.	16. 126.	
10.	8.	6.	8.	5.	9.	5.	5.	9.	7.	9. 71.	
11.	6.	6.	4.	6.	3.	1.	3.	1.	4.	5. 39.	
12.	1.	1.	2.	1.	1.	1.	2.	2.	1.	12.	
13.	1.	1.			1.		1.	1.	1.	6.	
14.											
15.											
16.											
	732	719	732.	700.	734.	698.	743.	722.	706.	737.	7210.

$$\int \frac{dx}{x} = 7212.99$$

How does the Gauss conjecture stand up to modern computing?

Recently, D. B. Staple computed that

$$\pi(10^{26}) = 1,699,246,750,872,437,141,327,603.$$

And Gauss would predict

$$\int_2^{10^{26}} \frac{dt}{\log t} = 1,699,246,750,872,592,073,361,408. \dots$$

The error is smaller than the square-root of the actual count!

This conjecture of Gauss may be viewed as saying it is appropriate to study the primes *statistically*.

It led to the **Riemann** Hypothesis (1859) (which is equivalent to the assertion that the logarithmic integral is within  $\sqrt{x} \log x$  of the true count). And to the prime number theorem (**Hadamard** & **de la Vallee Poussin** in 1896, **Erdős** & **Selberg** 1949) (which merely asserts that the ratio of the count to the integral tends to 1 as  $x \rightarrow \infty$ ).

More relevant to this talk, this *statistical* view of primes morphed into a *probabilistic* view. In 1923, **Hardy** and **Littlewood** conjectured that the density of twin primes near  $x$  is given asymptotically by  $c/(\log x)^2$ . That is,  $p$  and  $p + 2$  are “independent events” where the constant  $c \approx 1.32$  is a fudge factor to take into account the degree to which they’re *not* independent.



For example, the actual count of twin primes to  $10^{16}$  is

**10,304,195,697,298,**

computed by P. Sebah. The twin prime constant (fudge factor) is

$$c := 2 \prod_{p>2} \left( 1 - \frac{1}{(p-1)^2} \right) = 1.32032363169373915 \dots$$

And

$$c \int_2^{10^{16}} \frac{dt}{(\log t)^2} = \mathbf{10,304,192,554,496. \dots}$$

The error is again only about the square-root of the count! Despite this fantastic numerical agreement, we don't even know for sure that there are infinitely many twin primes.

Actually, in 1871, **Sylvester** came up with a similar heuristic for the number of representations of an even number as a sum of two primes (and so gave a heuristic for Goldbach's conjecture). Hardy and Littlewood returned to this in 1923, but revised Sylvester's constant. The Hardy–Littlewood constant seems to be the “right” one (following both the reasoning for the constant and numerical experiments).

In 1937, **Cramér** gave an explicitly probabilistic heuristic (citing the Borel–Cantelli lemma), that the length of the maximal gap between consecutive primes in  $[1, x]$  is  $\sim (\log x)^2$ . (In 1995, **Granville** revised Cramér's heuristic to take into account certain conspiracies that can deterministically occur among numbers divisible by a small prime, to get that the maximal prime gap is heuristically  $\sim c(\log x)^2$ , where  $c$  is perhaps  $2e^{-\gamma} \approx 1.1229$ .)

Also, the statistical/probabilistic view moved on beyond the primes themselves.

In 1917, **Hardy** and **Ramanujan** proved that the “normal” number of prime factors of an integer near  $x$  is  $\log \log x$ . (This means that for each fixed  $\epsilon > 0$ , the asymptotic density of the integers  $n$  with between  $(1 - \epsilon) \log \log n$  and  $(1 + \epsilon) \log \log n$  prime factors is 1.) Though clearly a statistical result, the proof was not.

In 1934, **Turán** gave a new and simple proof of the **Hardy–Ramanujan** theorem, that was based on the second-moment method in probability, but he didn’t realize that that is what he had done!

*“When writing Hardy first in 1934 on my proof of the Hardy–Ramanujan theorem, I did not know what Chebyshev’s inequality was and a fortiori on the central limit theorem. Erdős, to my best knowledge, was at that time not aware too. It was Mark Kac who wrote to me a few years later that he discovered when reading my proof in J. LMS that this is basically probability and so was his interest turned to this subject.”*

Letter of Paul Turán to Peter Elliott in 1976, quoted in Elliott’s  
“Probabilistic number theory, vol. II”

The distribution of “abundant” numbers (a topic going back to antiquity) was worked out in the 1920s and 1930s by **Schoenberg, Davenport** and others, culminating in the **Erdős–Wintner** theorem in 1939.

Also that year, we had the celebrated **Erdős–Kac** theorem on the Gaussian distribution of the number of prime factors of a number.

So was born “probabilistic number theory”, a vital part of analytic number theory.

But what of the “probabilistic method”, where one proves the existence of various strange things by showing that with a suitable probability distribution, there is a positive chance that they exist?

In 1931, **Sidon** wondered how dense a set of positive integers can be if no number has more than 1 intrinsic representation as a sum of two members of the set. (That is,  $a + b = n$  is considered as the same representation of  $n$  as  $b + a$ .) And what is the slowest growing function  $f(n)$  for a set where every number has at least one representation as a sum of two members, but not more than  $f(n)$  representations?

These problems became the subject of much research over the next 30 years, and some of the best theorems were proved via the probabilistic method:

**Erdős** (1954): One can take  $f(n)$  as  $c \log n$  for some  $c$ .

**Erdős** (1956): There's a set where every number  $n$  has between  $c_1 \log n$  and  $c_2 \log n$  representations as a sum of two elements.

Still unsolved: Is there a set and a constant  $c > 0$  such that every number  $n$  has  $\sim c \log n$  representations as a sum of two members of the set, as  $n \rightarrow \infty$ ?

In Sidon's original problem, he wondered about having at most one intrinsic representation. **Erdős** and **Rényi**, using the probabilistic method in 1960, showed that there is a fairly dense set where every number has a *bounded* number of representations as a sum of two members.

In a very recent paper, **Ford, Konyagin, Maynard, P, & Tao** used a variant of the probabilistic method to show there are anomalously long strings of composites in a polynomial sequence.

Let us shift gears to computational thoughts. If  $p$  is an odd prime, the function  $x^2 \bmod p$  is 2 : 1 for nonzero residues  $x$ , so there are exactly  $\frac{1}{2}(p - 1)$  nonzero squares mod  $p$  and exactly  $\frac{1}{2}(p - 1)$  non-squares mod  $p$ . Consider the algorithmic problem of finding one of these non-squares.

For example, for  $p = 3$ , 2 is a non-square. In fact, 2 works as a non-square for “half” of the primes, namely those that are 3 or 5 mod 8. For the prime 7, 3 is a non-square, and 3 works for the primes that are 5 or 7 mod 12. And so on.

This seems painlessly easy! But in fact, we do not have a deterministic polynomial time algorithm that produces a non-square for a given input prime  $p$ . (Assuming a generalized form of the Riemann Hypothesis allows us to prove that a certain simple algorithm runs in polynomial time.)



But in practice, no one is concerned with this, because we have a wonderful *random* algorithm that produces a non-square mod  $p$ . Namely, choose a random residue  $r \bmod p$  and check to see if it is a square or a non-square mod  $p$  (there is a simple polynomial-time check). The probability of success is  $\frac{1}{2}$ , and so the expected number of trials for success is 2.

This simple example is in fact closely tied to the fundamental problems of factoring polynomials over a finite field, and to primality testing.

For primality testing, we've long known of simple random algorithms that will quickly recognize composite numbers, leading us to strong conjectures that those not revealed as composite are prime.

Thirty-five years ago, **Adleman, P, & Rumely** found a deterministic primality test that is “nearly” polynomial time. And fifteen years ago, a true polynomial time primality test was found by **Agrawal, Kayal, & Saxena**.

These deterministic tests are not so computer practical; in practice we still rely on randomness and heuristics, even if we're searching for a proof of primality.

We also use probabilistic reasoning to construct deterministic algorithms.

An example is the *quadratic sieve* factoring algorithm that I found in the early 1980s. The method is almost completely heuristic, assuming numbers produced by a particular quadratic polynomial behave like random numbers of similar size.

(Shhh... No one should tell the large composites about this, they don't know we haven't rigorously proved that the quadratic sieve works, they get factored anyway!)

In fact, this state of affairs is largely true for all practical factoring algorithms, from the Pollard rho method, to the elliptic curve method, and the number field sieve. The elliptic curve method explicitly exploits randomness, but is still a heuristic method. The other algorithms, like the quadratic sieve, are deterministic, but with heuristic, probabilistic analyses.

So far we have considered the distribution of the primes, probabilistic number theory, the probabilistic method in number theory, and the role of randomness in number theoretic algorithms.

The probabilistic view also can help guide us in diophantine equations. For example, long before **Andrew Wiles** gave his celebrated proof of Fermat's Last Theorem (with help from **Richard Taylor**), we had a theorem of **Erdős** and **Ulam**.

They proved that if  $\mathcal{A}$  is a random set of natural numbers where  $a \in \mathcal{A}$  with probability  $\approx a^{-3/4}$ , then the number of triples  $a, b, c \in \mathcal{A}$  with  $a + b = c$  is almost surely bounded. Well the specific set of all powers higher than the third power forms a set  $\mathcal{A}$ , and the probability a random  $a \in \mathcal{A}$  is about  $a^{-3/4}$ . So this suggests that Fermat's Last Theorem is true with "probability 1".

There are a couple of caveats here. First, included in our specific set  $\mathcal{A}$  are the powers of 2 starting at  $2^4$ . And  $2^k + 2^k = 2^{k+1}$ , so there are infinitely many triples in the set  $a + b = c$ . These examples can be barred by assuming that  $a, b, c$  are coprime.

A second caveat, is that the same argument shows that with probability 1, a random set  $\mathcal{A}$ , where the probability of  $a \in \mathcal{A}$  is  $\approx a^{-2/3}$ , has infinitely many triples  $a, b, c$  with  $a + b + c$ . So Fermat's Last Theorem with exponent 3, is almost surely false!

But it's true, so it shows that the probabilistic view does not tell the whole story.

By the way, **Darmon** and **Granville** proved (using Faltings' theorem) that for any triple  $u, v, w$  with reciprocal sum  $\leq 1$ , there are at most finitely many coprime solutions to  $a^u + b^v = c^w$ .

Though Fermat's Last Theorem has been proved, and we have the Darmon–Granville theorem just above, what's still unknown is the **ABC Conjecture**. **Mochizuki** claims a proof, but it has not yet been accepted by the experts.

What is the ABC Conjecture, and why is it a conjecture?

For a positive integer  $n$ , let  $\text{rad}(n)$  denote the largest squarefree divisor of  $n$ ; that is,

$$\text{rad}(n) = \prod_{p|n} p.$$

**The ABC Conjecture:** For each  $\epsilon > 0$  there are at most finitely many coprime triples  $a, b, c$  with  $a + b = c$  and

$$c < \text{rad}(abc)^{1-\epsilon}.$$

It was posed by **Masser** and **Oesterlé** after **Mason** gave an elementary proof of the polynomial analogue.



We begin with a **Lemma**: *For each fixed  $\delta > 0$  and  $x$  sufficiently large, the number of integers  $n \leq x$  with  $\text{rad}(n) \leq y$  is  $\leq yx^\delta$ .*

Let  $i, j, k$  run over positive integers with  $i + j + k \leq (1 - \epsilon) \log x$ . For each  $i, j, k$  consider  $a, b \leq x$  and  $\frac{1}{2}x < c \leq x$  with

$$\text{rad}(a) \leq e^i, \quad \text{rad}(b) \leq e^j, \quad \text{rad}(c) \leq e^k.$$

Then  $\text{rad}(abc) \leq e^{i+j+k} \leq x^{1-\epsilon} < 2c^{1-\epsilon}$ . By the lemma, the number of choices for  $a$  is  $\leq e^i x^\delta$ , and similarly for  $b$  and  $c$ . So, the number of triples  $a, b, c$  is  $\leq e^{i+j+k} x^{3\delta} \leq x^{1-\epsilon+3\delta} = x^{1-\frac{1}{2}\epsilon}$ , assuming that  $\delta = \frac{1}{6}\epsilon$ . So the total # of triples:  $\leq x^{1-\frac{1}{2}\epsilon} \log^3 x$ .

Given  $a, b$ , the chance that a random  $c \in (\frac{1}{2}x, x]$  happens to be  $a + b$  is proportional to  $1/x$ , so letting  $a, b, c$  run, the chance we have an  $a, b, c$  triple is at most about  $x^{-\frac{1}{2}\epsilon} \log^3 x$ . Now let  $x$  run over powers of 2, and we get a convergent series.

The ABC Conjecture is hard to falsify, since it says there are at most finitely many counterexamples. Unlike with the Riemann Hypothesis or Fermat's Last Theorem, where even one counterexample can or could have destroyed the conjecture, this is not so for the ABC Conjecture.

In fact there are websites devoted to giving interesting “counterexamples”.

Take

$$2 + 3^{10} \cdot 109 = 23^5.$$

We have  $23^5 = 6,436,343$  and  $2 \cdot 3 \cdot 109 \cdot 23 = 15,042$ . See <http://www.math.unicaen.fr/~nitaj/abc.html>, a site maintained by **A. Nitaj**.

Another area where randomness has played a fundamental role: the **Cohen–Lenstra** heuristics.

Named after Henri Cohen and Hendrik Lenstra, these are a series of conjectures about the distribution of algebraic number fields (of given degree over the rationals), whose class groups have special properties. Basically their viewpoint is that groups should be weighted inversely by the size of their automorphism groups, but otherwise, assume randomness. They then produce concrete conjectures that can be tested statistically, and for the most part, they are looking quite good.

For example, statistically it is noticed that about 43% of class groups of imaginary quadratic field have 3-torsion, while the heuristic predicts 43.987%. And there seem to be about 76% of real quadratic fields with prime discriminant with class number 1, while the prediction is 75.446%.

Let me conclude with an idiosyncratic problem, one that **Erdős** once proclaimed as perhaps his favorite.

A finite set of integer residue classes is said to form a covering, if the union of the residue classes contains every integer.

A simple example:  $0 \pmod{1}$ .

Another simple example:  $0 \pmod{2}, 1 \pmod{2}$ .

Another:  $0 \pmod{3}, 1 \pmod{3}, 2 \pmod{3}$ .

To make this concept nontrivial, let's rule out the modulus 1, and let's also rule out repeated moduli.

A rule-abiding example:

$0 \bmod 2$ ,  $0 \bmod 3$ ,  $1 \bmod 4$ ,  $1 \bmod 6$ ,  $11 \bmod 12$

One can see this works by viewing each as 1 or more classes mod 12. Then  $0 \bmod 2$  hits the 6 even classes,  $0 \bmod 3$  hits 3 and 9,  $1 \bmod 4$  hits 1 and 5,  $1 \bmod 6$  hits 7, and  $11 \bmod 12$  hits 11.

**Erdős** conjectured in 1950 that there are coverings with distinct moduli where the least modulus is arbitrarily large.

The current record is held by **Nielsen** (2009) who found a covering with least modulus 40. The moduli only involve the primes to 107, but it has more than  $10^{50}$  of them!

This is nice, but where's the probability?

Let's consider a simple fact. If the moduli used are distinct primes, then they cannot cover, no matter what is chosen as representatives for the residue classes. Why?

Say the moduli are  $p_1, p_2, \dots, p_k$ , where these are distinct primes. Being in some residue class modulo one of these primes is an independent event from being in a class for another of them. In fact, the asymptotic density of the integers not covered will be exactly

$$\prod_{i=1}^k \left(1 - \frac{1}{p_i}\right),$$

which can be arbitrarily close to 0, but cannot be 0.

The exact same argument holds if the moduli  $m_1, m_2, \dots, m_k$  are merely pairwise coprime.

So the **Erdős** covering problem is very much one of extremal cases of *dependent* probabilities!

Some years ago I wondered what the maximal density one can cover using all of the integers in  $(x, 2x]$  as moduli. Would it be about

$$\sum_{m \in (x, 2x]} \frac{1}{m} \sim \log 2 \quad \text{or} \quad \prod_{m \in (x, 2x]} \left(1 - \frac{1}{m}\right) \sim \frac{1}{2}$$

or somewhere in between?



In the late 90s I discussed this problem with **Gang Yu**, a former student of mine at Georgia. He went off to U. South Carolina and began discussing the problem with **Michael Filaseta**.

Over some years a paper slowly developed of **Filaseta, Ford, Konyagin, P, & Yu** (2007). We proved among many other things that the moduli between  $x$  and  $2x$  behave asymptotically as if they're independent, that is, one cannot remove more than  $\frac{1}{2} + o(1)$  of the integers with them.

Our proof used a lemma that the referee pointed out to us resembles the **Lovász** local lemma, a central tool in the probabilistic method of combinatorics.

A few years ago at the **Erdős** centennial conference in Budapest, **Hough** announced his disproof of the **Erdős** covering conjecture! There is a minimal number  $B < 10^{16}$  such that any covering with distinct moduli must use a modulus at most  $B$ . We don't know what  $B$  is, but at least we know that  $B \in [40, 10^{16})$ .

Hough's proof used our version of the local lemma in a strong way. Using similar, but more involved methods, he and **Nielsen** just announced a proof that in any covering with distinct moduli, the moduli cannot all be coprime to 6. It's not known if there's a covering with all moduli odd. **Erdős** thought such a covering should exist, but **Selfridge** thought not.

There are many more links of number theory to probability, and I haven't even mentioned random number generators. Well, perhaps another time.

**Thank You**