TOPICS IN ANALYTIC NUMBER THEORY

LECTURES BY KANNAN SOUNDARARAJAN,
NOTES BY TONY FENG

CONTENTS

1. Overview 1
2. Distribution of \( L(1, \chi_d) \) 3

1. Overview

This class will focus on some recent work in analytic number theory.

1.1. Distribution of values of \( L \)-functions.

- Values at the edge of the critical strip: in particular \( L(1, \chi_d) \) where \( d \) is a fundamental discriminant.

- Values at the center of the critical strip: \( \zeta(\frac{1}{2} + it) \). Possible topics include:
  1. Selberg’s Theorem that \( \log |\zeta(\frac{1}{2} + it)| \) is approximately normal.
  2. Analogs for families of \( L \)-functions. (One reason that these results are hard is that they encode deep information such as non-vanishing of central values.)
  3. Connections with moment conjectures.

Conjecture 1.1.1.

\[
\int_0^T \left| \zeta\left(\frac{1}{2} + it\right) \right|^{2k} \, dt \sim c_k T (\log T)^k.
\]

We can show that

\[
\int_0^T \left| \zeta\left(\frac{1}{2} + it\right) \right|^{2k} \, dt \gg T (\log T)^k
\]

and conditionally on the Riemann Hypothesis

\[
\int_0^T \left| \zeta\left(\frac{1}{2} + it\right) \right|^{2k} \, dt \ll T (\log T)^k.
\]

Originally a proof was given with exponent \( k^2 + \epsilon \) by Soundararajan; the \( \epsilon \) was removed by Adam Harper.

(4) Refinements of moment conjectures.
1.2. **Sign changes in multiplicative functions (d’après Matomaki-Radziwill).** Let $\lambda(n) = (-1)^{\Omega(n)}$ be the Liouville function.

**Conjecture 1.2.1** (Chowla). Let $h_1, \ldots, h_k$ be distinct. Then

$$\sum_{n \leq x} \lambda(n+h_1) \ldots \lambda(n+h_k) = o(x), \text{ or even } o(x^{1/2+\varepsilon}).$$

**Example 1.2.2.** In particular, the conjecture predicts that every sign pattern $+ - + - \ldots \underbrace{- + \ldots +}_{k}$ appears $1/2^k$ of the time. We are very far from proving this. Hildebrand established that all 8 3-term patterns appear infinitely often; Matomaki-Radziwill-Tao proved that all 8 patterns appear with positive density.

The main theorem of Matomaki-Radziwill is:

**Theorem 1.2.3** (Matomaki-Radziwill). As $h \to \infty$,

$$\sum_{x \leq n \leq x+h} \lambda(n) = o(h)$$

for almost all $x \leq X$.

An analogous result for primes would say that almost all intervals of length $f(x) \log x$, where $f(x) \to \infty$ as $x \to \infty$, contain primes. However, this still seems out of reach.

**Remark 1.2.4.** I had assumed for a long time that results for the Möbius/Liouville functions should not be too difficult to replicate for primes, but this seems to be the case here.

Building on this work, Terence Tao showed

**Theorem 1.2.5** (Tao). We have

$$\sum_{W(x) \leq n \leq x} \frac{\lambda(n)\lambda(n+1)}{n} = o(\log W(x))$$

for any $W(x) \to \infty$.

Tao then used this to resolve the Erdős Discrepancy Problem:

**Conjecture 1.2.6.** For any $f: \mathbb{N} \to \{\pm 1\}$,

$$\sup_{d,K} \left| \sum_{k \leq K} f(dk) \right| \to \infty.$$

Roughly, this says that you cannot arrange signs $\pm 1$ such that the sum over all intervals is 0. One reason that this is delicate is because if you allow the value 0, then $\chi$ a real character is okay.

**Example 1.2.7.** If $\chi$ is a character mod 3, then we could try the multiplicative function

$$\tilde{\chi} = \begin{cases} \chi(p) & p \neq 3 \\ -1 & p = 3. \end{cases}$$
What happens if we sum this up to $x$? The sum vanishes over periods, so the only issue occurs in summing over powers of 3. There are only about $\log x$ powers of 3 at most $x$, so
\[ \sum_{n \leq x} \tilde{\chi}(n) \asymp \log x. \]

1.3. **Primes.** We might try to discuss the recent result of Maynard:

**Theorem 1.3.1 (Maynard).** There are infinitely many primes with no 7 in the decimal expansion.

This might be too hard, so we might replace this with the weaker result with decimal replaced by base $q$ for large enough $q$.

Another result in the spirit of decimal expansion of primes has to do with a refinement of the conjecture.

**Conjecture 1.3.2.** We have
\[ \# \{ p_n \leq x \mid p_n \equiv a \pmod{q}, p_{n+1} \equiv b \pmod{q} \} \sim \frac{\pi(x)}{\varphi(q)^2}. \]

We can then inquire into the next-order terms:
\[ \# \left\{ p_n \leq x \mid p_n \equiv a \pmod{q}, p_{n+1} \equiv b \pmod{q} \right\} \sim \frac{\pi(x)}{\varphi(q)^2} \left( 1 + c_1(q; a, b) \frac{\log \log x}{\log x} + c_2(q; a, b) \frac{1}{\log x} + \ldots \right) \]

**Theorem 1.3.3 (Lemke-Oliver, Soundararajan).** In the notation above, $c_1(q; a, b)$ is negative and big if $a \equiv b$, and positive and small if $a \neq b$.

2. **Distribution of $L(1, \chi_d)$**

For $d < -4$, the class number formula gives
\[ L(1, \chi_d) = \frac{\pi}{\sqrt{|d|}} h(d). \]
so understanding the distribution of $L(1, \chi_d)$ is equivalent to understanding the distribution of class numbers.

How can we bound this? We have
\[ L(1, \chi_d) = \sum_n \frac{\chi_d(n)}{n} = \prod_p \left( 1 - \frac{\chi_d(p)}{p} \right)^{-1}. \]

2.1. **The moment method.** One approach is to study the moments
\[ \sum_{d \leq X} L(1, \chi_d)^k. \]
You should think of $L(1, \chi_d)$ as basically a constant; it never gets very big or very small (of course we can’t prove this). So we should have
\[ \sum_{d \leq X} L(1, \chi_d)^k \sim XC(k). \]

We might ask for asymptotic formulas as $k \to \infty$ slowly. Then we might investigate the uniformity of $k \to \infty$ and also what happens when $k$ is a negative or even complex number. Then we might ask how to understand the distribution from the moments.
Question 2.1.1. Given the moments of a distribution, can you reconstruct the distribution? That is, do the moments (given that they exist) determine the distribution.

Let us first consider the situation of a compact space. Suppose you know
\[ \int_0^1 x^k d\mu(x) = m_k, \quad k = 0, 1, 2, \ldots \]
Then you know \( \mu \), because Weierstrass allows you to approximate any continuous function to arbitrary precision.

What about on \( \mathbb{R} \)? By Weierstrass again we can express \( \chi_{(a, \beta)} \) by polynomials nearby, or by Fourier series. The answer is that the moments do determine a distribution provided that they don’t grow too quickly. The proof should be by approximating indicator functions with polynomials again. The problem is that the polynomial will misbehave near \( \infty \). The enemy is that the moments will grow, meaning the distribution has mass at \( \infty \).

A precise discussion can be found in Feller’s book. A sufficient condition is something like
\[ \sum_k \left(m_{2k}^{1/2k}\right)^{-1} \text{ diverges.} \]
This holds, for instance, for the moments of a Gaussian. The \( k \)th moment is 0 for \( k \) odd and the number of ways of pairing off the \( k \) variables, which is something like \((k - 1) \cdot (k - 3) \cdot (k - 5) \cdot \ldots \). Then \( m_{2k}^{1/2k} \approx \sqrt{k} \), so the sum of the reciprocals does diverge.

2.2. Random distribution model. For fun, let’s think through the distribution of \( L(2, \chi_d) \).

Question 2.2.1. What does the distribution of \( L(2, \chi_d) \) look like?

We can model this as
\[ \sum \frac{X(n)}{n^2} \]
where \( X(n) \) is among 1, 0, \(-1\). What should its distribution be?

The function should be multiplicative, so it is not independent over \( n \), but it is independent on the prime numbers and then determined by multiplicativity. We should have \( X(p) = 0 \) approximately \( 1/p \) of the time; it turns out that the answer is about \( \frac{1}{p+1} \) (a random squarefree number wants to be a multiple of \( p \) even less than a random number, since the quotient needs also to be coprime to \( p \)). The remainder of the time, \( X(p) \) should have an equal chance to be 1 or \(-1\).

The distribution of \( L(2, \chi_d) \) will look rather discontinuous, since it is largely determined by the value at the first few primes. (This is even more apparent for \( L(100, \chi_d) \); where \( X(2) \) essentially determines the value). So the distribution should clump around the first few primes.

This is what makes \( L(1, \chi_d) \) so interesting. This is
\[ \prod_p \left( 1 - \frac{\chi_d(p)}{p} \right)^{-1} \]
and since $\sum 1/p$ diverges, the distribution is nice and continuous (not determined at an early stage).

We should think of this as being similar to

$$L(1, X) = \prod_p \left( 1 - \frac{X(p)}{p} \right)^{-1}$$

where $X$ is a random as discussed above. The main contribution is from

$$\sum \frac{X(p)}{p} \approx \log L(1, X).$$

**Question 2.2.2.** Does the sum $\sum \frac{X(p)}{p}$ converge almost surely?

We should so, because there should be a significant amount of cancellation among the terms. More generally,

**Question 2.2.3.** For what $\sigma$ does

$$\sum \frac{X(p)}{p^\sigma}$$

converge?

The sum can be approximated by an integral, to which we can apply integration by parts:

$$\sum \frac{X(p)}{p^\sigma} \approx \int \frac{X(t)}{t^\sigma} dt$$

$$\approx \int \frac{1}{t^{\sigma+1}} \left( \int X(u) du \right) dt$$

We expect this to converge for $\text{Re} \sigma > \frac{1}{2}$, because we expect

$$\sum_{n=1}^{N} X(n) \sim \sqrt{N}.$$

Fix $\sigma > 1/2$. Then we predict that

$$L(\sigma, \chi_d) \longleftrightarrow L(\sigma, X).$$

Therefore, we would like to show that

$$\sum_{|d| \leq X} L(1, \chi_d)^k \sim \#\{\text{fund. disc. } |d| \leq X\} \mathbb{E}(L(1, X)^k)$$

where here $\sum'$ means summing over only fundamental discriminants. We can also inquire as to the *uniformity* of $k$ with respect to $X$, and what happens with $k$ negative. (Positive $k$ measures the behavior of the largest values of $L(1, \chi_d)$ while negative $k$ measures the behavior of the negative values of $L(1, \chi_d)$.)
Before we do this, let’s think about something that could potentially go wrong for negative \( k \). If a particular value is very large or very small relative to the others, then it could dominate the sum. However, we can easily obtain an upper bound

\[
L(1, \chi_d) \ll C \log |d|.
\]

The reason is that we can first estimate

\[
L(1, \chi_d) \ll \sum_{n \leq |d|} \frac{\chi_d(n)}{n}
\]

since sums over periods of characters vanish, and then we have a slew of results available:

\[
\sum_{n \leq x} \chi_d(n) \ll \begin{cases} |d| & \text{trivial} \\ |d|^{1/2} \log |d| & \text{Polya-Vinogradov: if } x \geq \sqrt{d} \\ x^{1/2}|d|^\epsilon & \text{GRH} \\ x|d|^{-\delta} & \text{Burgess: if } x > |d|^{1/4+\epsilon} \end{cases}
\]

The upshot is that we have sufficiently good upper bounds to handle the positive moments.

In terms of lower bounds, the best result available is Siegel’s ineffective bound

\[
|d|^{-\epsilon} \ll \epsilon L(1, \chi_d).
\]

Therefore, we can only bound below the moments of negative \( k \) for \( |k| \leq \frac{1}{3} \). However, for statistical estimates we can overcome a few bad cases, using that Siegel zeros are very rare.

### 2.3. Computing moments

Let’s discuss some methods for computing the moments. As discussed, we can truncate the sum in the \( L \)-function to values at most \( |d| \), and write

\[
L(1, \chi_d) = \sum_{n \leq |d|} \frac{\chi_d(n)}{n} + \text{(small error)}.
\]

We break up the sum at some \( N \), which we imagine as a small power of \( x \):

\[
L(1, \chi_d) = \sum_{n \leq N} \frac{\chi_d(n)}{n} + \sum_{|d| > N} \frac{\chi_d(n)}{n} + \text{(small error)}.
\]

In estimating the moments, we have to deal with terms of the form

\[
\sum_{|d| \leq x} \left( \sum_{n=N}^{2N} \frac{\chi_d(n)}{n} \right)^2 = \sum_{n_1, n_2 > N} \sum_{|d| \leq x} \left( \frac{d}{n_1 n_2} \right).
\]

The symbol \( \frac{d}{n_1 n_2} \) is principal if \( n_1 n_2 = \Box \) (meaning is a square), and non-trivial otherwise. In the non-trivial case, the conductor is at most \( N^2 \), and we can use the “Pólya-Vinogradov bound”, with some modifications to account for the fact that we only sum
over fundamental discriminants, to bound the sum by \( \ll N^{1+\varepsilon} \). The sum over squares is roughly saying that \( n_1 \approx n_2 \), so the trivial bound for here is \( xN \). In total, we get

\[
\sum_{|d| \leq x} \left( \sum_{n=N}^{2N} \chi_d(n) \right)^2 \ll xN + N^3.
\]

(Compare with the trivial bound, which is \( xN^2 \).) If we substitute this into the formula for \( L(1, \chi_d) \) then we find that

\[
L(1, \chi_d) \ll \left( \sum_{|d| \leq x} \left| \sum_{n=N}^{2N} \frac{\chi_d(n)}{n} \right|^2 \right)^{1/2} \ll \frac{x}{N} + N.
\]

**Remark 2.3.1.** Here is one way of using the Pólya - Vinogradov bound in this case (since we're summing only over fundamental discriminants). We have

\[
\sum_{|d| \leq x} \chi(d) \approx \sum_{|d| \leq x} \sum_{\ell \mid d} \mu(\ell) \chi(d).
\]

Just restrict to \( d \) positive; the sum over negative \( d \) is more or less the same. This is

\[
\sum_{\ell} \sum_{d \leq x/\ell^2} \chi(d \ell^2).
\]

The inner guy is \( \ll q^{1/2} \) (forget log), and also you have the trivial bound \( x/\ell^2 \). If you use whichever is smaller, you get something that works.

In fact, we can even truncate the sum over \( |d|^{1/2} \) (by the argument discussed in Question 2.2.3). Therefore we can restrict our attention to \( N \leq \sqrt{x} \), so

\[
\frac{x}{N} + N \ll \frac{x}{N}.
\]

This is telling us that the tail sums have negligible contribution, so it suffices to consider the short sum. (We have bounded the variance, which lets us truncate the tail sum for all moments.)

If we grant this, then the rest of the proof is quite easy. Set \( N \) to be a small power of \( x \), say \( \leq x^{1/10} \). Then

\[
\sum_{|d| \leq x} \left( \sum_{n \leq N} \chi_d(n) \frac{d}{n} \right)^k \sum_{d=1}^{x} \frac{1}{n_1 \ldots n_k} \sum_{|d| \leq x} \left( \frac{d}{n_1 \ldots n_k} \right).
\]

We then use the same argument as before. If the product is not a square, we get cancellation. The main term is if \( n_1 \ldots n_k = \Box \). What should be the contribution from such terms?

Go back to the random model: \( \mathbb{E}(X(n_1) \ldots X(n_k)) = \mathbb{E}(X(n_1 \ldots n_k)) = 0 \) unless \( n_1 \ldots n_k = \Box \), and otherwise is the number of discriminants coprime to \( n_1 \ldots n_k \). So we get that the square case contributes

\[
\sum_{|d| \leq x} \left( \frac{d}{n_1 \ldots n_k} \right) = \# \{ |d| \leq x \} \mathbb{E}(X(n_1 \ldots n_k)).
\]
Therefore we expect that
\[
\sum_{|d| \leq X} \left( \sum_{n \leq N} \frac{\chi_d(n)}{n} \right)^k \sim \#\{|d| \leq x\} \mathbb{E} \left( \sum_{n \leq N} \frac{X(n)^k}{n} \right).
\]

Remark 2.3.2. Does this method work for \(L(\sigma, \chi_d)\) for \(\sigma > 1/2\)? The variance can still be estimated by
\[
\sum_{|d| \leq X} \left( \sum_{n=N}^{2N} \frac{\chi_d(n)}{n^\sigma} \right)^2 \ll \frac{xN}{N^{2\sigma}} + \frac{N^3}{N^{2\sigma}}
\]
and we get a saving for \(\sigma > 1/2\). We use this estimate to bound the tails for the sums for all \(k\). Then we estimate
\[
L(\sigma, \chi_d) = \sum_{n \leq N} \frac{\chi_d(n)}{n^\sigma} + \text{(small error)}
\]
imagining \(N \geq |d|^6\). We can then use the same method to compute:

We cannot actually compute the moments. Indeed, knowing a result
\[
\sum_{|d| \leq X} L(\sigma, \chi_d) \sim \#\{|d| \leq x\} \mathbb{E}(L(\sigma, X)^k)
\]
would imply the Lindelöf hypothesis.

If we are willing to assume GRH or GLH (the generalized Lindelöf hypothesis) then we can always prove that
\[
L(\sigma, \chi_d) = \sum_{n \leq N} \frac{\chi_d(n)}{n^\sigma} + \text{(small error)}.
\]
The argument is to study
\[
\frac{1}{2\pi i} \int_{(c)} L(\sigma + w, \chi_d) \Gamma(w) N^w d w
\]
We have the identity
\[
\frac{1}{2\pi i} \int_{(c)} \Gamma(w) x^w d w = \sum_{n=0}^\infty \frac{(-1)^n}{n!} x^{-n} = e^{-1/x}
\]
by shifting the contour and adding up the residues. Using this above,
\[
\frac{1}{2\pi i} \int_{(c)} L(\sigma + w, \chi_d) \Gamma(w) N^w d w = \sum_{n \leq N} \frac{\chi_d(n)}{n^\sigma} \frac{1}{2\pi i} \int_{(c)} \left( \frac{N}{n} \right)^w \Gamma(w) d w
\]
\[
= \sum_{n \leq N} \frac{\chi_d(n)}{n^\sigma} e^{-n/N}.
\]
To exploit Lindelof/Riemann, we move $c$ to $1/2$. Since $\Re w = 1/2 - \sigma < 0$, we encounter one pole of $\Gamma$ at 0. The residue there gives $L(\sigma, \chi_d)$, and what's left is
\[
\frac{1}{2\pi i} \int_{\frac{1}{2} - i\infty}^{\frac{1}{2} + i\infty} L(\sigma + w, \chi_d) \Gamma(w) N^w \, dw.
\]
GRH/GLH would give $L(\sigma + w, \chi_d) \ll (|d| (1 + |w|))^\varepsilon$, and $\Gamma$ decreases exponentially, and then we get a saving from $N$, so this is $O(|d|^\varepsilon N^{1/2 - \sigma})$.

The conclusion is that at least on GRH,
\[
\sum_{|d| \leq X} L(\sigma, \chi_d)^k \sim \#\{|d| \leq X\} \mathbb{E}(L(\sigma, X)^k) \quad \text{for } \sigma > 1/2.
\]

Remark 2.3.3. You can see that everything fails for $\sigma = \frac{1}{2}$, as we get no saving and the expectation also doesn't converge.

2.4. Largest and smallest values. We discuss other methods for studying the distribution of $L(1, \chi_d)$.

Question 2.4.1. Here is a “toy problem”. Suppose we consider a sequence of “coin tosses” $X(n) = \pm 1$ each with probability $1/2$. The sum
\[
\sum_{n=1}^{\infty} \frac{X(n)}{n}
\]
What is the chance that this sum is at least 10?

The biggest contributor to the chance is the outcome $X(1) = X(2) = \ldots = X(\lfloor e^{10} \rfloor) = 1$, whose probability is very small: about $2^{-e^{10}}$.

What is the “probability” that $L(1, \chi_d) > e^\tau \tau$? We know that
\[
\prod_{p \leq P} (1 - 1/p)^{-1} = e^\tau \log P
\]
so we want $\log P \approx \tau$, or $P \approx e^\tau$. Then the probability that $L(1, \chi_d) > e^\tau \tau$ is
\[
2^{-\pi(P)} = \exp(-C \frac{e^\tau}{\tau}).
\]
Thus $L(1, X)$ has double exponentially decaying tail:
\[
\mathbb{P}(L(1, X) > e^\tau \tau) = \exp\left(-C \frac{e^\tau}{\tau} (1 + o(1))\right).
\]
How small can $L(1, \chi_d)$ get? You should then consider
\[
\prod_{p \leq P} \left(1 + \frac{1}{p}\right)^{-1} = \frac{C'}{\log P}.
\]
This also gives doubly exponentially decay.

Using this we can try to predict the largest size of $L(1, \chi_d)$ and the smallest size of $L(1, \chi_d)$. We choose $\tau$ so that the expectation of attaining a certain value is just right. We earlier took $P = e^\tau$. What is the chance that for all $p \leq P$ we have $\left( \frac{d}{p} \right) = 1$. The chance of attaining the value $\tau$ is $2^{-\pi(P)} \cdot X$, so
\[
\pi(P) \sim \log x \\
P \sim \log x \cdot (\log \log x).
\]
Since $e^\tau / \tau = P$ we predict

\[
\text{largest value} = e^\tau(\log \log x + \log \log \log x + c) \\
\text{smallest value} = e^\tau(\log \log x + \log \log \log x + c)^{-1}
\]

Remark 2.4.2. How much does one get from GRH? Soundararajan-Granville showed that

\[
L(1, \chi_d) \prod_{p \leq [\log |d|]^{2}} \left( 1 - \frac{\chi_d(p)}{p} \right)^{-1} \leq 2e^\gamma(\log \log x)
\]
so this is off by a factor of 2.