Fractional Moments and Distribution

Mollifiers, Moments, and the Zeta Function — Lecture 2 of 3

Recap

- In the last lecture we introduced mollifiers, i.e. Dirichlet polynomials

$$M(s) = \sum_{n \le X} \frac{a(n)}{n^s}$$

with the property that $M(s) \approx 1/\zeta(s)$ most of the time (in the precise sense that $|\zeta(s)M(s)| \ll 1$ most of the time).

- The famous Selberg mollifier

$$M(s) = \sum_{n \le X} \frac{\mu(n)}{n^s} \cdot \Big(1 - \frac{\log n}{\log X}\Big)$$

is obtained by finding the coefficients $a(n)$ that minimize $\int_T^{2T} |\zeta(\frac{1}{2} + it) \sum_{n \le X} a(n) n^{-1/2-it}|^2 \, dt$.

Drawback of the Selberg mollifier

- One important drawback of the Selberg mollifier is that its *inverse* is not close to ζ. In no reasonable sense do we have

$$\Big(\sum_{n \leq X} \frac{\mu(n)}{n^s} \cdot \Big(1 - \frac{\log n}{\log X} \Big) \Big)^{-1} \approx \sum_{n \leq X} \frac{1}{n^s}.$$

- This is a significant drawback if we want to study the *distributional* aspects of the Riemann ζ-function.
- **Goal of this lecture:** salvage the idea of Bohr by constructing Dirichlet polynomials that behave like Euler products *on typical sets*, and apply this to fractional moments.

Moments as a proxy for distribution

- A proxy for the distributional behavior of the Riemann ζ-function are the moments,

$$\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k} dt.$$

- We can think of this as the moment generating function of $\log|\zeta|$ since the moment can be re-written as

$$\frac{1}{T}\int_T^{2T} \exp(2k\log|\zeta(\tfrac{1}{2}+it)|)dt.$$

- It is conjectured that the above is $\sim c_k \exp(k^2 \log\log T)$.
- This strongly suggests that $\log|\zeta(\frac{1}{2}+it)|$ should have Gaussian behavior.

Theorem 1 (Selberg)

$$\frac{1}{T}\mathrm{meas}_{T\le t\le 2T}\Big\{\frac{\log|\zeta(\frac{1}{2}+it)|}{\sqrt{\frac{1}{2}\log\log T}} > V\Big\} \sim \frac{1}{\sqrt{2\pi}}\int_V^\infty e^{-u^2/2}\,du.$$

- Conjectured to remain true all the way to $V = o(\sqrt{\log\log T})$.
- Work of Arguin–Bailey in this direction.

What is known about moments?

- We are able to compute $\int_T^{2T} |\zeta(\frac{1}{2}+it)|^{2k} dt$ for $k = 0, 1, 2$ and that's it. This is done by a direct computation.
- Can we say anything about fractional moments? Yes.
- In this lecture we will show how to obtain

$$\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k} dt \asymp T(\log T)^{k^2}$$

for simplicity we will just focus on the upper bound part.

First attempt: Cauchy–Schwarz (wrong exponent)

- For $0 \le k \le 1$, Cauchy–Schwarz gives

$$\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k} dt \le T\Big(\frac{1}{T}\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^2 dt\Big)^k$$

- Since $\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^2 dt \ll T\log T$, this gives

$$\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k} dt \ll T(\log T)^k$$

but the actual upper bound has a power of k^2.

Introducing exponential weights

- How to improve it? Let's pretend that we could use

$$\exp\left(\beta\Re\sum_{p\le X}\frac{1}{p^{1/2+it}}\right)$$

  as a mollifier.

- Recall the **random model**: replace p^{-it} by independent random variables $X(p)$ uniform on the unit circle. Then $\mathcal{P}(s) := \Re\sum_{p\le X}p^{-s}$ plays the role of a sum of independent random variables.

- Write

$$\int_T^{2T}|\zeta|^{2k}\,dt = \int_T^{2T}|\zeta|^{2k}\exp\Big(-\gamma\mathcal{P}\Big)\exp\Big(\gamma\mathcal{P}\Big)dt.$$

Applying Hölder

- Hölder with exponents $1/k$ and $1/(1-k)$ gives

$$\Big(\int_T^{2T} |\zeta|^2 \exp\Big(-\frac{\gamma}{k}\mathcal{P}\Big)dt\Big)^k \cdot \Big(\int_T^{2T} \exp\Big(\frac{\gamma}{1-k}\cdot\mathcal{P}\Big)dt\Big)^{1-k}.$$

Optimizing γ

- In the random model $\mathcal{P}$ is approximately Gaussian with variance $\frac{1}{2}\log\log T$ and $|\zeta|^2 \approx e^{2\mathcal{P}}$. Substituting the random model prediction $\mathbb{E}[e^{\alpha\mathcal{P}}] \sim (\log T)^{\alpha^2/4}$ into the Hölder bound, the exponent from Hölder is:

$$f(\gamma) = \frac{k(2-\gamma/k)^2}{4} + \frac{(1-k)(\gamma/(1-k))^2}{4}.$$

The correct γ

- Set $\alpha = \gamma/k$ and minimize: $f'(\alpha) = 0$ gives

$$(2-\alpha)(1-k) = k\alpha \implies \alpha = 2(1-k),$$

i.e.

$$\boxed{\gamma = 2k(1-k).}$$

- Check: $\gamma/k = 2(1-k)$, $\gamma/(1-k) = 2k$.
- Exponents: $k \cdot (2k)^2/4 = k^3$ and $(1-k) \cdot (2k)^2/4 = k^2(1-k)$.
- Total: $k^3 + k^2 - k^3 = k^2$. ✓

The main problem

- The main problem is that we are not able to compute the resulting expressions. The exponential $\exp(\beta\mathcal{P})$ is *not* a short Dirichlet polynomial.
- What if instead we first aimed to create a short Dirichlet polynomial that behaves "most of the time" like the Euler product?
- This is the idea behind Bohr's approach from Lecture 1, now carried out on *typical sets* rather than everywhere.
- For this we need a little bit of probabilistic intuition.

Typical values of the Euler product

- Consider

$$\exp\left(\beta\Re\sum_{p\le X}\frac{X(p)}{\sqrt{p}}\right)$$

where $X(p)$ are independent random variables uniform on the unit circle. Then the bulk of the values come from values of $X(p)$ for which

$$\left|\Re\sum_{p\le X}\frac{X(p)}{\sqrt{p}}\right|\le\left(\frac{\beta}{2}+\varepsilon\right)\sum_{p\le X}\frac{1}{p}.$$

Taylor expansion on the typical set

- On the typical set, we expand the exponential in a Taylor expansion,

$$\exp\left(\beta\Re\sum_{p\le X}\frac{X(p)}{\sqrt{p}}\right)=\sum_{j\ge 0}\frac{1}{j!}\left(\beta\Re\sum_{p\le X}\frac{X(p)}{\sqrt{p}}\right)^j.$$

- The terms past $j\ge\beta\log\log X$ give a negligible contribution. Thus on the typical set we can approximate the Euler product by a Dirichlet polynomial of length

$$X^{\beta\log\log X}.$$

- For this to be short enough ($\le T^{1/2}$) we need $X\le\exp\left(\frac{\log T}{(\log\log T)^2}\right)$ say.

Improving by splitting into two scales

- Can we improve it? The idea is to set $X_1 = T^{1/(\log\log\log T)^2}$ and look at

$$\exp\left(\beta \Re \sum_{X \le p \le X_1} \frac{X(p)}{\sqrt{p}}\right).$$

- The same principle applies but we can cut off earlier! The bulk of the values come from X for which

$$\left|\Re \sum_{X \le p \le X_1} \frac{X(p)}{\sqrt{p}}\right| \le \left(\frac{\beta}{2} + \varepsilon\right) \sum_{X \le p \le X_1} \frac{1}{p}.$$

Second scale: shorter polynomial

- We can truncate again at $j \leq \beta \log\log\log T$ this time. The resulting Dirichlet polynomial has length
$$X_1^{\beta \log\log\log T},$$
which means that we can take
$$X_1 = \exp\Big(\frac{\log T}{(\log\log\log T)^2}\Big).$$

- Overall: an expression for Euler products up to $T^{1/(\log\log\log T)^2}$ on typical sets. One can continue in this way, adding further scales.

Setup: two scales of primes

We now make this rigorous. Choose two even integers:

$$\ell_1 = 2\lceil(\log\log T)^2\rceil, \qquad \ell_2 = 2\lceil(\log\log\log T)^2\rceil.$$

Set thresholds and cutoffs:

$$L_1 = \ell_1/e^2, \qquad L_2 = \ell_2/e^2,$$
$$Y = T^{1/(2\ell_1)}, \qquad Z = T^{1/(2\ell_2)}.$$

Two scales of primes:

- $P_1 = \{p \le Y\}$, $\quad \mathcal{P}_1(t) = \Re \sum_{p\in P_1} p^{-1/2-it}$, $\quad$ variance $\sigma_1^2 \approx \frac{1}{2}\log\log T$.
- $P_2 = \{Y < p \le Z\}$, $\quad \mathcal{P}_2(t) = \Re \sum_{p\in P_2} p^{-1/2-it}$, $\quad$ variance $\sigma_2^2 \asymp \log\log\log T$.

Why these choices?

- On $\mathcal{T}_j = \{|\mathcal{P}_j| \le L_j\}$, we expand $e^{\alpha \mathcal{P}_j}$ in a Taylor polynomial of degree ℓ_j. The approximation is good when $|\mathcal{P}_j| \le \ell_j / e^2 = L_j$.
- Expanding $\mathcal{P}_j^{\ell_j}$ gives a Dirichlet polynomial of length $\le Y^{\ell_1} = T^{1/2}$ (resp. $Z^{\ell_2} = T^{1/2}$). Short enough.
- $L_1/\sigma_1 \sim (\log\log T)^2/\sqrt{\log\log T} = (\log\log T)^{3/2} \to \infty$, so $\mathcal{E}_1$ is small. Similarly $L_2/\sigma_2 \to \infty$.

Definition of the level sets

$$\mathcal{T}_1 = \{t \in [T, 2T] : |\mathcal{P}_1(t)| \le L_1\}, \qquad \mathcal{E}_1 = [T, 2T] \setminus \mathcal{T}_1,$$
$$\mathcal{T}_2 = \{t \in [T, 2T] : |\mathcal{P}_2(t)| \le L_2\}, \qquad \mathcal{E}_2 = [T, 2T] \setminus \mathcal{T}_2.$$

Start with the entire moment. Split:

$$\int_T^{2T} |\zeta|^{2k}\,dt = \int_{\mathcal{T}_1} |\zeta|^{2k}\,dt + \int_{\mathcal{E}_1} |\zeta|^{2k}\,dt.$$

Bound $\mathcal{E}_1$ using Hölder ($0 \le k \le 1$):

$$\int_{\mathcal{E}_1} |\zeta|^{2k} \le \mathrm{meas}(\mathcal{E}_1)^{1-k} \Big(\int_T^{2T} |\zeta|^2\,dt \Big)^k.$$

Chebyshev for $\mathcal{E}_1$

- Chebyshev with the $2m$-th moment of $\mathcal{P}_1$: for any even integer $2m$,

$$\mathrm{meas}(\mathcal{E}_1) \le \frac{1}{L_1^{2m}} \int_T^{2T} |\mathcal{P}_1(t)|^{2m}\, dt.$$

- The moments of $\mathcal{P}_1$ are computable (expand, use MVT for Dirichlet polynomials):

$$\int_T^{2T} |\mathcal{P}_1|^{2m}\, dt \sim T \cdot \sigma_1^{2m} \cdot \frac{(2m)!}{2^m m!}.$$

- By Stirling, optimizing in m:

$$\mathrm{meas}(\mathcal{E}_1) \ll T \cdot \exp\big(-c\, L_1^2/\sigma_1^2\big).$$

- Since $L_1/\sigma_1 \to \infty$ this is negligible, giving $\int_{\mathcal{E}_1} |\zeta|^{2k} \ll T$.

Step 2: On $\mathcal{T}_1$, introduce the exponential

- On $\mathcal{T}_1$ we can introduce $e^{\gamma \mathcal{P}_1} \cdot e^{-\gamma \mathcal{P}_1}$:

$$\int_{\mathcal{T}_1} |\zeta|^{2k} = \int_{\mathcal{T}_1} |\zeta|^{2k} \, e^{-\gamma \mathcal{P}_1} \cdot e^{\gamma \mathcal{P}_1} \, dt.$$

- Then split into $\mathcal{T}_2$ and $\mathcal{E}_2$:

$$= \int_{\mathcal{T}_1 \cap \mathcal{T}_2} |\zeta|^{2k} \, e^{-\gamma \mathcal{P}_1} \cdot e^{\gamma \mathcal{P}_1} \, dt + \int_{\mathcal{T}_1 \cap \mathcal{E}_2} |\zeta|^{2k} \, e^{-\gamma \mathcal{P}_1} \cdot e^{\gamma \mathcal{P}_1} \, dt.$$

- On $\mathcal{T}_1 \cap \mathcal{T}_2$: introduce the additional exponential $e^{\gamma \mathcal{P}_2} \cdot e^{-\gamma \mathcal{P}_2}$ and apply Hölder:

$$\int_{\mathcal{T}_1 \cap \mathcal{T}_2} |\zeta|^{2k} \, e^{-\gamma(\mathcal{P}_1 + \mathcal{P}_2)} \cdot e^{\gamma(\mathcal{P}_1 + \mathcal{P}_2)} \, dt$$

$$\leq \left(\int |\zeta|^2 \, e^{-(\gamma/k)(\mathcal{P}_1 + \mathcal{P}_2)} \, dt \right)^k \left(\int e^{(\gamma/(1-k))(\mathcal{P}_1 + \mathcal{P}_2)} \, dt \right)^{1-k}.$$

Step 3 (cont'd): Expanding and evaluating

- Since we are on both $\mathcal{T}_1$ and $\mathcal{T}_2$, $|\mathcal{P}_j| \le L_j = \ell_j/e^2$ on both scales. So we can expand:

$$e^{\alpha \mathcal{P}_j} \approx E_{\ell_j}(\alpha \mathcal{P}_j) := \sum_{m=0}^{\ell_j} \frac{(\alpha \mathcal{P}_j)^m}{m!}.$$

- Since ℓ_j is even, E_{ℓ_j} is non-negative. So we can extend from $\mathcal{T}_1 \cap \mathcal{T}_2$ to all of $[T, 2T]$ (upper bound).
- The resulting Dirichlet polynomials have length $\le Y^{\ell_1}, Z^{\ell_2} \le T^{1/2}$. Evaluate using the mean value theorem for Dirichlet polynomials.
- Recall $\gamma = 2k(1-k)$ from the earlier optimization. This gives $\ll T(\log T)^{k^2}$.

Step 4: Error term — $\mathcal{T}_1 \cap \mathcal{E}_2$ (Hölder)

- On $\mathcal{T}_1 \cap \mathcal{E}_2$: apply Hölder keeping the exponentials as they are:

$$\int_{\mathcal{T}_1 \cap \mathcal{E}_2} |\zeta|^{2k} \, e^{-\gamma \mathcal{P}_1} \cdot e^{\gamma \mathcal{P}_1} \, dt$$

$$\leq \Big(\int |\zeta|^2 \, e^{-\gamma \mathcal{P}_1 / k} \, dt \Big)^k \Big(\int_{\mathcal{T}_1 \cap \mathcal{E}_2} e^{\gamma \mathcal{P}_1 / (1-k)} \, dt \Big)^{1-k}.$$

- The first factor is a mollified second moment, handled as before.

Step 4 (cont'd): Expand on $\mathcal{T}_1$, Chebyshev on $\mathcal{E}_2$

- For the second factor: we are on $\mathcal{T}_1$ so we write the exponential as a square and expand, $e^{\gamma\mathcal{P}_1/(1-k)} = (e^{\gamma\mathcal{P}_1/(2(1-k))})^2 \approx |D_1(t)|^2$, where D_1 is the Taylor expansion of $e^{\gamma\mathcal{P}_1/(2(1-k))}$. We are also on $\mathcal{E}_2$, so add a Chebyshev bound:

$$\int_{\mathcal{T}_1\cap\mathcal{E}_2} |D_1|^2\,dt \;\le\; \int_{\mathcal{E}_2} |D_1|^2\,dt \;\le\; \frac{1}{L_2^{2m}}\int_T^{2T} |D_1|^2\,|\mathcal{P}_2|^{2m}\,dt.$$

- D_1 depends on primes in P_1, $\mathcal{P}_2$ depends on primes in P_2, these are **disjoint**, so:

$$\int |D_1|^2\,|\mathcal{P}_2|^{2m}\,dt \;\approx\; \frac{1}{T}\Big(\int |D_1|^2\,dt\Big)\Big(\int |\mathcal{P}_2|^{2m}\,dt\Big).$$

- This decoupling + Chebyshev makes the error negligible.

Putting it all together

$$\int_T^{2T} |\zeta(\tfrac{1}{2}+it)|^{2k}\,dt = \underbrace{\ll T(\log T)^{k^2}}_{\mathcal{T}_1\cap\mathcal{T}_2} + \underbrace{\text{negligible}}_{\mathcal{E}_1} + \underbrace{\text{negligible}}_{\mathcal{T}_1\cap\mathcal{E}_2}.$$

- One can continue iterating to push the primes closer to T, capturing more of the variance. The structure repeats at each new scale.
- Next time: Selberg's central limit theorem. The multi-scale Taylor expansion reappears, and taking $k \to 0$ in the fractional moment heuristic leads naturally to the CLT.