Plan of the talk:

- A little bit of number theory notation/background
- Erdős-Kac Central Limit Theorem
- Random multiplicative functions
- Some ideas about prime number races/3(\frac{1}{2}+i\epsilon)

(0) Number Theory background

Let \( w(n) := \# \{ \text{distinct prime factors of } n \} \) \hspace{1cm} (e.g. \( w(12) = w(2^2 \cdot 3) = 2 \))
Let \( \mathcal{P}(n) := \text{largest prime factor of } n \) \hspace{1cm} (\geq 2)
Let \( \mu(n) := \begin{cases} 0 & \text{if } n \text{ is divisible by a non-trivial square} \\ 1 & \text{if } n \text{ is squarefree and } w(n) \text{ is even} \\ -1 & \text{if } n \text{ is squarefree and } w(n) \text{ is odd} \end{cases} \)

(e.g. \( \mu(2) = \mu(3) = \mu(\text{prime}) = -1, \mu(4) = 0, \mu(6) = 1 \))

Theorem: (Prime Number Theorem, Hadamard, de la Vallée Poussin, 1896).

\[ \# \{ p \leq x : p \text{ prime} \} = \left(1+o(1)\right) \frac{x}{\log x} \quad \text{as } x \to \infty. \]

Notation: we write \( f(x) \ll g(x) \) to mean \( f(x) \leq C g(x) \) (i.e. the same as \( f(x) \leq C g(x) \)).
Erdős–Kac Central Limit Theorem

We can write $w(n) = \sum_{p \leq n} S_p(n)$, where $S_p(n) = \begin{cases} 1 & \text{if } p \mid n \\ 0 & \text{otherwise} \end{cases}$.

If we randomly choose an integer $n \leq x$, then $P(S_p = 1) = \frac{1}{\log x} \cdot \frac{\pi(x)}{\log x}$.

If $n \leq x$, then $P(S_p = 1) = \frac{1}{\log x} \cdot \frac{\pi(x)}{\log x}$.

And $P(S_p = 1 \text{ and } S_q = 1) = \frac{1}{\log x} \cdot \frac{\pi(x)}{\log x}$.

So $w(n)$ is a sum of indicators that behave "roughly independently" — so maybe there is a CLT here?

Theorem: (Erdős–Kac, 1939–40)

If $n \leq x$ is chosen uniformly at random, then

$$w(n) \sim \frac{\log \log x}{\log x}$$

as $x \to \infty$.

This is now very classical, but in my opinion it is still an open problem to give a "nice" proof of the sharp rate of convergence.

* the correct rate is a Berry–Essen type rate $\frac{1}{\log \log x}$

(Rényi–Turán proved this, but proofs is difficult complex and
there is a very simple qualitative proof using moments, by Billingsley, 1969 [Erdős-Kac also used moments, but it was much messier].

I used the Chen-Stein method to give a total variation bound \( O(\frac{\log \log x}{\log x}) \) between a truncated version of \( \omega(n) \) (don't count very large prime factors) and Poisson — this leads to a rate \( O(\frac{\log \log \log x}{\log \log x}) \) in Erdős-Kac.

Nice open problem: prove such a total variation approximation (maybe with a weaker rate) without truncating \( \omega(n) \) ? This would show that \( \omega(n) \) is odd and even asymptotically \( \frac{1}{2} \) the time, which is equivalent to PNT.

7) Random multiplicative functions

This is a random model introduced by Wintner, 1940, as a heuristic for the
Moebius function \( \mu(n) \) [and can also think of it as a heuristic for the characteristic function \( \chi_{(\mathbb{Z}/2\mathbb{Z})^*} \) that are real-valued, with \( \gamma \) very large]

Model: Let \( \{f(p)\}_p \) prime be a sequence of independent Rademacher RVs

\( f(p) = 1 \) if \( n \) has a non-trivial square divisor, \( f(n) = -1 \) if \( f(n) = 0 \) o/w.
Limit Theorems

**Question**: Is it true that \( \frac{\sum_{n \leq x} f(n)}{\sqrt{\sum_{n \leq x} f(n)^2}} \xrightarrow{d} N(0,1) \) as \( x \to \infty \)?

**Answer**: Note that \( \sum_{n \leq x, \omega(n) \leq k} f(n) \) does satisfy the CLT (sum of independent Rademachers). Hough, 2011, proved using moments that \( \sum_{n \leq x, \omega(n) = k} f(n) \) satisfies CLT, provided \( k = o(\log log x) \) (i.e., just fewer than average prime factors).

I proved 2013 using martingales that CLT holds provided \( k = o(\log log x) \).

(We can write \( \sum_{n \leq x, \omega(n) = k} f(n) = \sum_{n \leq x} f(n) \) a martingale decomposition w.r.t. the prime revealing filtration).

This reduces matters to a fourth moment calculation.

I also proved that \( \sum_{n \leq x} f(n) / \sqrt{\sum_{n \leq x} f(n)^2} \xrightarrow{d} N(0,1) \) as \( x \to \infty \). (by a conditional argument).

Chatterjee & Soundararajan, 2011: proved that CLT holds for \( \sum_{n \leq x, \gamma(n) \leq k} f(n) \), if

The proof uses Chatterjee's `new method` \( C x^{1/2} \log x < y = o(\log x) \).
Open problems: What is the correct range of $y$ in Chatterjee & Soundararajan's result?

II. Does $\sum_{n \leq x} f(n)$ have a limit distribution, and if so what is it?

Size of $\sum_{n \leq x} f(n)$

We know that $\sum_{n \leq x} f(n) = O_e \left( \sqrt{x} \left( \log \log x \right)^{5/2+\epsilon} \right)$ a.s. for any $\epsilon > 0$ (Lau, Tenenbaum & Wu, 2013) \leftarrow Uses Bombieri–Vaughan inequality.

$\sum_{n \leq x} f(n) \neq O \left( \frac{x}{(\log \log x)^{3/2+\epsilon}} \right)$ (H., 2013) \leftarrow Uses Gaussian processes.

Open problem: What is the true size of the fluctuations of $\sum_{n \leq x} f(n)$?

In particular, is it true that $\mathbb{E} \left| \sum_{n \leq x} f(n) \right| \geq c \log x$? (Conjecture: yes, but other people might disagree.)

3) Other topics.

Prime number race: Let $q \in \mathbb{N}$, and let $a_0, \ldots, a_r$ be some of the coprime residue classes mod $q$. (possibly all)
Let \( n(x, q, a_i) := \# \{ \text{primes } p \leq x : p \equiv a_i \mod q \} \).

Question: As \( x \to \infty \), how often do we have
\[
 n(x, q, a_1) \geq n(x, q, a_2) \geq \ldots \geq n(x, q, a_r) ?
\] "a, w.r.t. \( a_2 \) comes second..."

Assuming GRH & Li

Rubinstein & Samak showed that as \( x \to \infty \), the distribution of
\( (n(x, q, a_1), \ldots, n(x, q, a_r)) \) has a limit (when suitably centred and
rescaled). This limit is the same as the distribution of certain
weighted sums of independent RVs (weights depend on \( a_i \)).

So to answer the Question, we can try to approximate \( \xi_q(a_1, \ldots, a_r) \)
by a multivariate Gaussian distribution, and then use tools like Gaussian
comparison inequalities. Can be done using Stein's