Central limit theorem for random multiplicative functions

Sourav Chatterjee
(Courant Institute)

joint work with Kannan Soundararajan (Stanford)

Multiplicative functions

- Many of the functions of interest to number theorists are multiplicative. That is, they satisfy $f(mn) = f(m)f(n)$ for all coprime natural numbers m and n.
- Examples: the Möbius function $\mu(n)$, the function n^{it} for a real number t, Dirichlet characters $\chi(n)$.
- Often one is interested in the behavior of partial sums $\sum_{n\leq x} f(n)$ of such multiplicative functions.
- For the prototypical examples mentioned above it is a difficult problem to obtain a good understanding of such partial sums.
- A guiding principle that has emerged is that partial sums of specific multiplicative functions (e.g. characters or the Möbius function) behave like partial sums of random multiplicative functions.
- For example, this viewpoint is explored in the context of finding large character sums in Granville and Soundararajan (2001).

Random multiplicative functions

- Values of the multiplicative function X at primes are chosen independently at random, and the values at squarefree numbers are built out of the values at primes by the multiplicative property.
- For example, $X(2)$, $X(3)$ and $X(5)$ are independent random variables, while $X(30) = X(2)X(3)X(5)$.
- Define $X(n) = 0$ for n that is not squarefree (as for the Möbius function), retaining the multiplicative property.
- In this talk, for each prime p, $X(p)$ is either $+1$ or -1 with equal probability. Thus, $X(n) \in \{-1, 0, 1\}$ for all n.

Summatory behavior of random multiplicative functions

- Let $M(x) := \sum_{n \le x} X(n)$.
- Easy to show: $\mathbb{E}(X(n)) = 0$, $\mathrm{Var}(X(n)) \le 1$ (with equality if n is squarefree), and $X(n), X(m)$ are uncorrelated for $n \ne m$.
- Follows that for every x,

$$\mathrm{Var}(M(x)) = \#\{\text{squarefree numbers} \le x\}.$$

- Since the density of squarefree numbers is $6/\pi^2$, this implies that for any fixed x, the typical fluctuation of $M(x)$ is of order $\sqrt{x}$.
- But in a given realization of the random function, there may be anomalous x, just by chance, which exhibit larger fluctuations.
- Halasz (1982) showed that there are constants c and d such that with probability 1, $|M(x)|$ is bounded by the function

$$c\sqrt{x}\exp(d\sqrt{\log\log x \log\log\log x}).$$

(Also proved a nearly matching lower bound.)

Distribution of partial sums?

- Halasz's result can be viewed as Law of Iterated Logarithm for random multiplicative functions.
- Naturally raises the question of proving a central limit theorem.
- We know $\mathbb{E}(M(x)) = 0$, $\mathbb{E}(M(x)^2) \sim \frac{6}{\pi^2}x$ for large x.
- If one can show that for all k

$$\lim_{x\to\infty} \frac{\mathbb{E}(M(x)^k)}{(6x/\pi^2)^{k/2}} \stackrel{?}{=} \mathbb{E}(Z^k), \qquad (*)$$

  where Z is a standard Gaussian random variable, this would prove the CLT. (This is called the method of moments.)
- But equation (*) is not true! The limit is ∞ for all even k above a threshold.
- However, this does not disprove the CLT. Central limit theorems can hold even without moments converging.

Recent results

- Hough (2008) showed that for each fixed k, if $M_k(x)$ is the sum of $X(n)$ over all $n \le x$ that have k prime factors, then $M_k(x)$ satisfies a CLT. Proof by method of moments.
- Harper (2009) showed that $M_k(x)$ satisfies a CLT even if k is allowed to grow like $(1-\delta)\log\log x$ for any $\delta > 0$. Proof by martingales.
- However, Harper (2009) showed that if k grows like $(1+\delta)\log\log x$, then CLT is no longer true!
- In particular, the partial sum $M(x)$ does not satisfy a Gaussian CLT. (Recall: $M(x) = \sum_{n\le x} X(n)$.)
- Recall that most numbers $\le x$ have approximately $\log\log x$ prime factors. Harper's result gives an interesting dichotomy.
- It seems from simulations that $M(x)/\sqrt{x}$ has a limiting distribution as $x \to \infty$. But we do not know what it is.

Next question: sums in small intervals

- Sometimes, sums of multiplicative functions in small intervals like $[x, x+y]$, where $y \ll x$, are of interest.
- Can we analyze the behavior of $M(x,y) := \sum_{x<n\le x+y} X(n)$, where X is our random multiplicative function?
- Unless y grows very slowly (slower than $\log x$), the high moments of
$$\frac{M(x,y)}{\sqrt{\mathrm{Var}(M(x,y))}}$$
blow up as $x \to \infty$ and $y \to \infty$, rendering the method of moments useless for proving a CLT, just as for $M(x)$.
- Question: Does the CLT hold if y grows like x^α for some $\alpha < 1$?

The main result

The following theorem shows that the CLT for $M(x, y)$ holds as long as y grows slower than $x/(\log x \log\log x)$.

Theorem

Let X be our random multiplicative function and

$$M(x, y) := \sum_{x<n\leq x+y} X(n).$$

Let $S(x, y)$ be the number of squarefree integers in $(x, x+y]$. If $x \to \infty$ and $y \to \infty$ such that $y = o(x/(\log x \log\log x))$, then

$$\frac{M(x, y)}{\sqrt{S(x, y)}} \xrightarrow{\text{distribution}} \text{standard Gaussian},$$

provided $S(x, y)/y$ remains bounded away from zero.

Remark: The last condition is satisfied if y grows faster than $x^{1/5+\epsilon}$, by a result of Filaseta and Trifonov (1992).

Large and small primes

- Note: if we change the value of $X(p)$ for some small prime p (e.g. $p = 2$), $M(x, y)$ must undergo a large change. On the other hand, central limit theorems arise mainly as a 'sum of many small independent contributions'. If one $X(p)$ contributes so much, how can we expect a CLT? This is the main reason why CLT fails for $M(x)$.
- This is taken care of by dividing the set of primes into 'small' and 'large' primes, and then conditioning on the small primes.
- Let x, y be as in the statement of the theorem, and $\delta = y/x$.
- Let $z := \frac{1}{2}\log(1/\delta)$.
- Divide the primes below $2x$ into the large (that is $> z$) and small (that is $\leq z$) primes, denoted by $\mathcal{L}$ and $\mathcal{S}$.
- Let $\mathcal{F}$ be the sigma-algebra generated by $X(p)$ for all $p \in \mathcal{S}$, and denote the conditional expectation given $\mathcal{F}$ by $\mathbb{E}^{\mathcal{F}}$.

Small primes do not matter

- Recall: $S(x, y) =$ number of squarefree integers in $(x, x + y]$.
- The key step in the proof is to show that the conditional distribution of $M(x, y)$ given the sigma-algebra $\mathcal{F}$ is approximately Gaussian with mean 0 and variance $S(x, y)$, irrespective of the values of $(X(p))_{p \in \mathcal{S}}$.
- A basic probabilistic fact is that if the conditional distribution of a random variable Y given a sigma-algebra $\mathcal{F}$ is a non-random distribution F, then the unconditional distribution of Y is again F.
- This fact, combined with the above claim about the conditional distribution, implies that the unconditional distribution of $M(x, y)$ is approximately Gaussian with mean 0 and variance $S(x, y)$.

First indication of the irrelevance of small primes

Recall: $\mathcal{F}$ is the sigma-algebra generated by the values of X at the small primes.

Lemma

Irrespective of the values of $X(p)$ for $p \in \mathcal{S}$, we have $\mathbb{E}^{\mathcal{F}}(M(x,y)) = 0$ and $\mathbb{E}^{\mathcal{F}}(M(x,y)^2) = S(x,y)$,

- To prove $\mathbb{E}^{\mathcal{F}}(M(x,y)) = 0$, we only need observe that any $n \in (x, x+y]$ must have a prime factor in $\mathcal{L}$.
- This is easy, because the product of all primes in $\mathcal{S}$ is less than x.
- To prove $\mathbb{E}^{\mathcal{F}}(M(x,y)^2) = S(x,y)$, it suffices to prove that $X(n)$ and $X(n')$ are uncorrelated even after conditioning on $\mathcal{F}$, for any $n \neq n'$ in $(x, x+y]$.
- Again, this is easy because if $n \neq n'$, there must exist distinct $p, p' \in \mathcal{L}$ such that $p|n$ and $p'|n'$.

The conditional CLT

- The previous lemma shows that the first and second moments of $M(x, y)$, conditional on the values of X at the small primes, do not actually depend on these values.
- This needs to be extended to show that the full distribution of $M(x, y)$, conditional on the values of X at the small primes, is approximately independent of these values.
- Program: Fix any set of values of $X(p)$ for $p \in \mathcal{S}$. Then $M(x, y)$ is simply a function of $X(p), p \in \mathcal{L}$.
- Perturbing any $X(p)$ for $p \in \mathcal{L}$ creates only a relatively small perturbation in $M(x, y)$.

An abstract central limit theorem

- Suppose $X = (X_1, \ldots, X_n)$ and $X' = (X'_1, \ldots, X'_n)$ are i.i.d. random vectors with independent components.
- Let $W = f(X)$ be a function of X with mean 0 and var 1.
- For each $A \subseteq \{1, \ldots, n\}$, define the vector X^A as: $X_i^A = X'_i$ if $i \in A$, and $X_i^A = X_i$ if $i \notin A$.
- Let $\Delta_j f(X) := f(X) - f(X^j)$.
- Define

$$T := \frac{1}{2} \sum_A \frac{1}{\binom{n}{|A|}(n - |A|)} \sum_{j \notin A} \Delta_j f(X) \Delta_j f(X^A).$$

Theorem (C., 2008)

Let $Z \sim N(0,1)$. Then for any Lipschitz function ϕ,

$$|\mathbb{E}\phi(W) - \mathbb{E}\phi(Z)| \le \sqrt{\mathrm{Var}(T)} + \frac{1}{2} \sum_{j=1}^n \mathbb{E}|\Delta_j f(X)|^3.$$

Simplest example

- Suppose $f(X) = n^{-1/2}\sum_{i=1}^n X_i$. Then a simple computation gives
$$T = \frac{1}{2n}\sum_{j=1}^n (X_j - X_j')^2.$$
Thus, $\mathrm{Var}(T) = O(n^{-1})$.
- Also,
$$\sum_{j=1}^n \mathbb{E}|\Delta_j f(X)|^3 = n^{-3/2}\sum_{j=1}^n \mathbb{E}|X_j - X_j'|^3 = O(n^{-1/2}).$$
- Combining, we get an $O(n^{-1/2})$ error bound.

Applying the abstract CLT to our problem

- Fixing $X(p)$ for $p \in \mathcal{S}$, $M(x, y)$ can be considered as a function of the independent r.v. $X(p), p \in \mathcal{L}$.
- Computing T for this function is simple. Getting suitable estimates for $\mathrm{Var}(T)$ involves expectations of sums of products like $X(n_1)X(n_2)X(n_3)X(n_4)$. (Requires some results from number theory.)
- The cubic remainder term is small because perturbation of $X(p)$ for large primes produces a small effect on $M(x, y)$.
- Combination of the above gives the desired CLT for $M(x, y)$ (conditional on the values of X at small primes).
- Unconditional CLT is derived by the principle mentioned before.

Brief sketch of the proof of the abstract CLT

- First, recall the notation:
- $X = (X_1, \ldots, X_n)$ and $X' = (X_1', \ldots, X_n')$ are i.i.d. random vectors with independent components. $W = f(X)$ is a function of X with mean 0 and var 1. For each $A \subseteq \{1, \ldots, n\}$, the vector X^A is defined as: $X_i^A = X_i'$ if $i \in A$, and $X_i^A = X_i$ if $i \notin A$. $\Delta_j f(X) := f(X) - f(X^j)$. Finally, T is defined as

$$T := \frac{1}{2} \sum_A \frac{1}{\binom{n}{|A|}(n - |A|)} \sum_{j \notin A} \Delta_j f(X) \Delta_j f(X^A).$$

- Thus, for any absolutely continuous function ψ,

$$\mathbb{E}(\psi'(W)T) = \frac{1}{2} \sum_A \frac{1}{\binom{n}{|A|}(n - |A|)} \sum_{j \notin A} \mathbb{E}(\psi'(W) \Delta_j f(X) \Delta_j f(X^A)).$$

- Next step: simplify $\mathbb{E}(\psi'(W) \Delta_j f(X) \Delta_j f(X^A))$.

Proof sketch continued

- If $\Delta_j f(X)$ is small, then with $g = \psi \circ f$, we have the approximate chain rule $\Delta_j g(X) \approx \psi'(f(X))\Delta_j f(X) = \psi'(W)\Delta_j f(X)$.
- Thus,

$$\mathbb{E}(\psi'(W)\Delta_j f(X)\Delta_j f(X^A)) \approx \mathbb{E}(\Delta_j g(X)\Delta_j f(X^A))$$
$$= \mathbb{E}(g(X)\Delta_j f(X^A)) - \mathbb{E}(g(X^j)\Delta_j f(X^A)).$$

- Swapping the roles of X_j, X'_j inside the second expectation, we get $\mathbb{E}(g(X^j)\Delta_j f(X^A)) = -\mathbb{E}(g(X)\Delta_j f(X^A))$. Combined with the previous step, this gives

$$\mathbb{E}(\psi'(W)\Delta_j f(X)\Delta_j f(X^A)) \approx 2\mathbb{E}(\psi(W)\Delta_j f(X^A))$$

- Combining all steps, we have

$$\mathbb{E}(\psi'(W)T) \approx \mathbb{E}\biggl(\psi(W)\sum_A \frac{1}{\binom{n}{|A|}(n-|A|)}\sum_{j\notin A}\Delta_j f(X^A)\biggr).$$

Proof sketch continued

- A simple algebraic verification shows that

$$\sum_A \frac{1}{\binom{n}{|A|}(n-|A|)} \sum_{j \notin A} \Delta_j f(X^A) = f(X) - f(X').$$

- Recalling that $W = f(X)$, this gives

$$\begin{aligned} \mathbb{E}(\psi'(W)T) &\approx \mathbb{E}[\psi(W)(f(X) - f(X'))] \\ &= \mathbb{E}(\psi(W)W) - \mathbb{E}(\psi(W))\mathbb{E}(f(X')) \\ &= \mathbb{E}(\psi(W)W), \text{ since } \mathbb{E}(f(X')) = \mathbb{E}(W) = 0. \end{aligned}$$

- Exact equality holds for $\psi(u) = u$, which gives $\mathbb{E}(T) = \mathbb{E}(W^2) = 1$.
- Thus, if $\mathrm{Var}(T)$ is tiny, then "we can replace T by 1", and get

$$\mathbb{E}(\psi(W)W) \approx \mathbb{E}(\psi'(W)).$$

Finishing off with Stein's method

- We have shown that for any ψ, $\mathbb{E}(\psi(W)W) \approx \mathbb{E}(\psi'(W))$.
- Given a Lipschitz ϕ, produce a function ψ that solves the o.d.e.

$$\psi'(x) - x\psi(x) = \phi(x) - \mathbb{E}\phi(Z).$$

- Use basic o.d.e. theory to show that ψ is sufficiently well-behaved.
- Then

$$\mathbb{E}\phi(W) - \mathbb{E}\phi(Z) = \mathbb{E}(\psi'(W) - \psi(W)W) \approx 0.$$

- The above idea is the foundation of Stein's method of distributional approximation. This completes the proof.