Finding the distribution of random multiplicative functions in short intervals

Adam J Harper
University of Warwick

June 2025

This talk is based on joint work with K. Soundararajan and Max Xu.

Plan of the talk:

- Introduction, definitions, review of previous work
- A new result
- Why should you believe the new result?

Introduction

In 1944, Wintner constructed a random real sequence whose values have a multiplicativity property, to model the Möbius function.

Definition 1

Let $(f(p))_{p \text{ prime}}$ *be independent Rademacher random variables (i.e. taking values* ± 1 *with probability* $1/2$ *each). We define a* Rademacher random multiplicative function *by setting* $f(n) := \prod_{p|n} f(p)$ *for all squarefree* n*, and* $f(n) = 0$ *when* n *is not squarefree.*

Thus $f(nm) = f(n)f(m)$ provided n, m are coprime (as for μ).
And $f(n) = \mu(n) = 0$ if n has any non-trivial square divisors.
On squarefree n, $f(n)$ and $\mu(n)$ take values ± 1.

To model functions like $n \mapsto n^{-it}$ or complex Dirichlet characters $\chi(n)$, we also consider:

Definition 2

Let $(f(p))_{p \text{ prime}}$ be independent Steinhaus random variables (i.e. distributed uniformly on the unit circle $\{|z| = 1\}$). We define a Steinhaus random multiplicative function *by setting $f(n) := \prod_{p^a || n} f(p)^a$ for all n, where $p^a || n$ means that p^a is the highest power of p that divides n.*

In the Steinhaus case, we have $f(mn) = f(m)f(n)$ for all m, n (as for $n \mapsto n^{-it}$ or $\chi(n)$).

Key probabilistic point: The values $f(n)$ are **not** all independent. For example, if we know $f(2)$ and $f(3)$ then $f(6) = f(2)f(3)$ is entirely determined.

Questions

The most obvious question to ask would seem to be:
" **What is the distribution of** $\frac{1}{\sqrt{x}} \sum_{n \leq x} f(n)$**, as** $x \to \infty$ **?** "

In view of the classical Central Limit Theorem (*although that doesn't apply here*), one might expect the distribution to be roughly (real or complex) Gaussian.

But in fact, we have:

Theorem 0 (H., 2020)

If $f(n)$ is a Rademacher or Steinhaus random multiplicative function, then uniformly for all large x and real $0 \leq q \leq 1$ we have

$$\mathbb{E}|\sum_{n \leq x} f(n)|^{2q} \asymp \left(\frac{x}{1 + (1-q)\sqrt{\log \log x}} \right)^q .$$

In particular, $\mathbb{E}|\sum_{n \leq x} f(n)| \asymp \frac{\sqrt{x}}{(\log \log x)^{1/4}}$. "Better than squareroot cancellation"

This implies that $\frac{1}{\sqrt{x}} \mathbb{E}|\sum_{n \leq x} f(n)| \ll \frac{1}{(\log \log x)^{1/4}} \to 0$ as $x \to \infty$.
So $\frac{1}{\sqrt{x}} \sum_{n \leq x} f(n)$ **converges** (in probability) **to zero** as $x \to \infty$.

For the full sum, to get non-degenerate behaviour one should actually study $\frac{(\log\log x)^{1/4}}{\sqrt{x}}\sum_{n\leq x} f(n)$ instead of $\frac{1}{\sqrt{x}}\sum_{n\leq x} f(n)$. ("*Non-trivial renormalisation*")

The problem of finding the limiting distribution of this (or even whether one exists) remains open, although there has been important recent progress by Gorodetsky–Wong and by S. Hardy.

" **What is the distribution of** $\frac{1}{\sqrt{y}} \sum_{x<n\le x+y} f(n)$ **?** "

Here we think of the regime where $y = y(x)$ satisfies $y \to \infty$ as $x \to \infty$, but $y = o(x)$ (growing but short intervals).

Important Note: In the classical central limit theorem (for sums of IID random variables ϵ_n), there is *no difference* in distribution between $\sum_{n\le y} \epsilon_n$ and $\sum_{x<n\le x+y} \epsilon_n$.
But in the random multiplicative setting, the multiplicativity could interact with the short interval of summation in a non-trivial way.

Known results

In the positive direction, it was known that $\frac{1}{\sqrt{y}} \sum_{x<n\le x+y} f(n)$ **does have** a (non-degenerate) Gaussian limiting distribution provided:

- $x^{1/5} \log x \ll y = o(\frac{x}{\log x})$ for f Rademacher; (Chatterjee–Soundararajan, 2012)
- $y \to \infty$ and $y \le \frac{x}{\log^{2\log 2-1+\epsilon} x}$ for f Steinhaus; $x^{1/5} \log x \ll y \le \frac{x}{\log^{2\log 2-1+\epsilon} x}$ for f Rademacher. (Soundararajan–Xu, 2023)

In the negative direction:

- using the moment bound of H., we see

$$\begin{aligned}\frac{1}{\sqrt{y}}\mathbb{E}|\sum_{x<n\leq x+y} f(n)| &\leq \frac{1}{\sqrt{y}}(\mathbb{E}|\sum_{n\leq x+y} f(n)| + \mathbb{E}|\sum_{n\leq x} f(n)|) \\ &\ll \frac{1}{\sqrt{y}}\frac{\sqrt{x}}{(\log\log x)^{1/4}} \\ &\to 0 \quad \text{as } x\to\infty,\end{aligned}$$

  provided that $\frac{(x/y)}{\sqrt{\log\log x}} \to 0$.

- looking inside the proof of the moment bound, Caich (2024) shows that $\frac{1}{\sqrt{y}}\mathbb{E}|\sum_{x<n\leq x+y} f(n)| \to 0$ provided that $\frac{\log(x/y)}{\sqrt{\log\log x}} \to 0$.

When $\frac{\log(x/y)}{\sqrt{\log\log x}} \gg 1$, Caich shows instead that $\frac{1}{\sqrt{y}}\mathbb{E}|\sum_{x<n\leq x+y} f(n)| \asymp 1$.

Our new result

Theorem 1 (H.–Soundararajan–Xu, in preparation)

If $f(n)$ is a Steinhaus random multiplicative function and $\frac{x}{\log^{0.4} x} \le y = o(x)$ (say), then the following is true.

- *There exists a deterministic quantity $V(x,y)$ such that*

$$\frac{1}{\sqrt{V(x,y)}} \sum_{x<n\le x+y} f(n) \xrightarrow{d} \textit{standard complex Gaussian} \quad \text{as } x \to \infty.$$

- *$V(x,y)$ satisfies $V(x,y) \asymp y \min\{1, \frac{\log(x/y)}{\sqrt{\log\log x}}\}$, and also $V(x,y) \sim y$ as $\frac{\log(x/y)}{\sqrt{\log\log x}} \to \infty$.*

We also have the much easier (given known results):

Theorem 2 (H.–Soundararajan–Xu, in preparation)

If $f(n)$ is a Steinhaus or Rademacher random multiplicative function, and $\delta > 0$ is small, then the following is true. If λ is large enough in terms of δ; and x is large enough in terms of λ; and $\delta x \leq y \leq x$; then

$$\mathbb{P}(|\sum_{x<n\leq x+y} f(n)| > \lambda \frac{\sqrt{y}}{(\log \log x)^{1/4}}) \gg \frac{1}{\lambda^C}.$$

Combined with results discussed earlier, this implies there is **no way** to renormalise $\sum_{x<n\leq x+y} f(n)$ (with $y \geq \delta x$ and δ fixed) to possibly get a non-degenerate Gaussian limit.

Remarks:

- Theorem 2 essentially just follows from the low moment bounds and Hölder's inequality, exploiting the blow-up as one approaches the second moment.
- An analogue of Theorem 1 should also hold in the Rademacher case, we may or may not include this depending on our energy levels!
- You should think that one gets convergence to a Gaussian whenever $y = o(x)$, but *the rate of convergence becomes worse and worse the slower that x/y grows*. Once $x/y \asymp 1$, one can no longer have convergence to a (non-degenerate) Gaussian.

Why should you believe Theorem 1?

For the full sum $\frac{1}{\sqrt{x}} \sum_{n \leq x} f(n)$, it is fairly accurate to think the distribution is like a Gaussian with *random variance*

$$\approx \frac{1}{\log x} \int_{-1}^{1} |F(1/2 + it)|^2 dt,$$

where $F(s)$ is the random Euler product corresponding to $f(n)$ on x-smooth numbers (e.g. $F(s) = \prod_{p \leq x} (1 - \frac{f(p)}{p^s})^{-1}$ in the Steinhaus case).

- It turns out that the typical size of $\frac{1}{\log x} \int_{-1}^{1} |F(1/2+it)|^2 dt$ is $\asymp \frac{1}{\sqrt{\log\log x}}$, hence the "better than squareroot cancellation" for the full sum.
- It also turns out that the distribution of $\frac{1}{\log x} \int_{-1}^{1} |F(1/2+it)|^2 dt$ has *heavy tails*, hence the overall distribution of $\frac{(\log\log x)^{1/4}}{\sqrt{x}} \sum_{n \leq x} f(n)$ is *not Gaussian*.

For the short interval sum $\frac{1}{\sqrt{y}} \sum_{x<n\leq x+y} f(n)$, it is fairly accurate to think the distribution is like a Gaussian with random variance

$$\approx \frac{1}{(x/y)\log x} \int_{-(x/y)}^{(x/y)} |F(1/2+it)|^2 dt.$$

(This can be established using martingale theory + rather a lot of work!)

- It turns out that the typical size of $\frac{1}{(x/y)\log x}\int_{-(x/y)}^{(x/y)} |F(1/2+it)|^2 dt$ is $\asymp \min\{1, \frac{\log(x/y)}{\sqrt{\log\log x}}\}$. Hence the moment bounds of Caich, and the size of $V(x,y)$ in our theorem.
- **Crucial Point:** $\int_{-(x/y)}^{(x/y)} |F(1/2+it)|^2 dt \approx \sum_{|n|\le x/y} \int_{n-1/2}^{n+1/2} |F(1/2+it)|^2 dt$, where the subintegrals $\int_{n-1/2}^{n+1/2} |F(1/2+it)|^2 dt$ are *roughly independent* of one another.
  So the random variance $\frac{1}{(x/y)\log x}\int_{-(x/y)}^{(x/y)} |F(1/2+it)|^2 dt$ *concentrates around something deterministic* (cf Law of Large Numbers).