On twisted sums of random multiplicative functions (Warwick)

Ofir Gorodetsky

1 Motivation, history

Let $\alpha_S\colon \mathbb{N} \to S^1$ be Steinahus (completely multiplicative), $\alpha_S\colon \mathbb{N} \to \{-1,0,1\}$ be Rademacher.
Original motivation: α_R models μ. RH is $\sum_{n\le x}\mu(n) \ll x^{1/2+\varepsilon}$. Wintner (1944) proved $\sum_{n\le x}\alpha_R(n) \ll x^{1/2+\varepsilon}$ holds almost surely **for all** x.
It is expected that α_R misses some properties of μ. E.g. $\sum_{n\le x}\mu(n) = O(\sqrt{x}(\log\log\log x)^{5/4})$ is expected (Gonek) while Harper proved that if $V(x)\to\infty$ then $\sum_{n\le x}\alpha_R(n) \ge \sqrt{x}(\log\log x)^{1/4}/V(x)$ for infinitely many x a.s. On the other hand, α_S is the large-T limit of $(n^{it})_{t\le T}$ and the large-q limit of (χ mod q), e.g.

$$\sum_{n\le x} n^{it} \xrightarrow{d} \sum_{n\le x}\alpha_S(n)$$

as $T\to\infty$ (since $(p^{it})_{p\le x} \xrightarrow{d} (\alpha_S(p))_{p\le x}$; here t is uniform in $[0,T]$). And we certainly want to understand such sums – they build up ζ.
Significant work on moments of $\sum_{n\le x}\alpha_S(n)$ and law of iterated logarithm (Halász, Erdős, Lau-Tenenbaum-Wu, Basquin, Harper, Caich). Focus of this talk: distribution.
The distribution of $\sum_{n\le x}\alpha_S(n)$, appropriately normalized is still not known, but very recently a precise conjecture was stated (2024). Harper's work (2017) shows that normalizing by standard deviation gives trivial limiting distribution (Helson), and that 'true' normalization is by $\sqrt{x}/(\log\log x)^{1/4}$. This relates to critical multiplicative chaos. Lots of activity on variants:

$$\sum_{n\le x}\alpha_S(n)f(n).$$

E.g. very recently (2025), Hardy proved a version of the conjecture where $f(n) = \mathbf{1}_{P(n)>\sqrt{n}}$. Most works, until Hardy's, focused on much 'sparser' f, where correct normalization turns out to be **standard deviation**, and limiting distribution was **standard Gaussian**:

1. $f(n) = \mathbf{1}_{\omega(n)=k}$: Hough, Harper. $k = o(\log\log x)$. The sum has length $x/\log^{1+o(1)} x$.
2. $f(n) = \mathbf{1}_{[x,x+H]}$: Chatterjee–Sound ($H = o(x/\log x)$), Sound–Xu ($H \le x/(\log x)^{2\log 2-1+\varepsilon}$), Pandey–Wang–Xu ($H \ll_A x/\log^A x$ for every $A>0$), Harper–Sound–Xu.
3. $f(n) = \mathbf{1}_{Q(\mathbb{Z})}$: Najnudel, Klurman–Shkredov–Xu, Wang–Xu, Chinis–Shala.

None of these f are multiplicative.

2 Work with Wong

Motivated by works of Najnudel–Paquette–Sim, Mo Dick Wong and I considered

$$S_x := \Big(\sum_{n\le x}|f(n)|^2\Big)^{-1/2}\sum_{n\le x}\alpha_S(n)f(n)$$

for multiplicative f, with $|f(p)|^2$ equal to $\theta\in(0,1/2)$ on average; recently extended to $|f(p)|^2$ equal to $\theta\in(0,1)$ on average, inspired by ideas in Najnudel–Paquetee–Simm–Vu. So cannot cover $f\equiv 1$, but can take e.g. indicator of sums of two squares. In fact, f does not have to be bounded, $f(p)$ can infinitely often be as large as $p^{1/2+o(1)}$ ($\sum_p |f(p)|^3\log^C p/p^{3/2}<\infty$). If f takes the values 0 and 1 only, its support is $\asymp x(\log x)^{\theta-1}$

What is the limiting distribution we found? $G \cdot \sqrt{V}$ for standard complex Gaussian G independent of V, and $V = 1/(2\pi)\int_{\mathbb{R}} |1/2+it|^{-2} m_\infty(dt)$. Here m_∞ is a random measure, translation-invariant, constructed by a limit from a sequence of measures involving α and f. We use the notation G for standard complex Gaussian throughout. This talk will not be about m_∞ (see Mo Dick's talk for that), but about **how one connects the random sum with the random measure**.

One approach for this is through moments. It is not relevant here because the moments of V explode: $\mathbb{E}V^p$ is finite iff $p < 1/\theta$, which forces $\mathbb{E}|S_x|^{2p}$ to diverge for $p \ge 1/\theta$ (recall that if $X_n \xrightarrow{d} X$ in distribution then $\liminf \mathbb{E}|X_n|^{2p} \ge \mathbb{E}|X|^{2p}$ by Fatou's lemma.)

Next approach is martingale CLT, introduced into this area by Harper (2010). Why martingales?

$$S_x = \sum_{p \le x} Z_{p,x}$$

where

$$Z_{p,x} := (\sum_{n \le x} |f(n)|^2)^{-1/2} \sum_{n \le x,\, P(n)=p} \alpha_S(n) f(n).$$

If we define filtrations $\mathcal{F}_{p^-} = \sigma((\alpha(q) : q < p)$ then $\mathbb{E}[Z_{p,x} \mid \mathcal{F}_{p^-}] = 0$ (essentially conditioning of values of $\alpha(q)$ for $q < p$.) Let

$$V_x = \sum_{p \le x} |Z_{p,x}|^2.$$

Informally, $S_x \approx G\sqrt{V_x}$. Formally, **McLeish CLT** makes this formal in a special case: if $\mathbb{E}V_x \to 1$, $\mathbb{E}V_x^2 \to 1$, and $\sum_{p \le x} \mathbb{E}|Z_{p,x}|^4 \to 0$, then $S_x \xrightarrow{d} G$. Note that these conditions imply $V_x \xrightarrow{p} 1$. First condition is trivial: $\mathbb{E}V_x = 1$ regardless of f. (Alternative to $\mathbb{E}V_x^2 \to 1$ is $\mathbb{E}|S_x|^4 \to 2$; see Sound–Xu.)

This CLT is not relevant if limiting distribution is not Gaussian, so we turn to the general form of martingale CLT. We define

$$BP_x := \sum_{p \le x} \mathbb{E}[|Z_{p,x}|^2 \mid \mathcal{F}_{p^-}].$$

If $BP_x \xrightarrow{p} V$, $\sum_{p \le x} \mathbb{E}|Z_{p,x}|^4 \to 0$ and $\sum_{p \le x} \mathbb{E}[Z_{p,x}^2 \mid \mathcal{F}_{p^-}] \xrightarrow{d} 0$, then $S_x \xrightarrow{d} G\sqrt{V}$. Rest of the talk will be about $BP_x \xrightarrow{p} V$ – other conditions are 'trivial' to verify, *even in critical case* (e.g. $\sum_{p \le x} \mathbb{E}[Z_{p,x}^2 \mid \mathcal{F}_{p^-}] \equiv 0$).

3 Two approaches to bracket process

It is convenient to assume that f is supported on squarefrees. Moreover, we shall only talk about Steinhaus (we put $\alpha = \alpha_S$). We suppose $\sum_{p \le x} |f(p)|^2 \sim \theta \mathrm{Li}(x)$. First step is to understand how BP_x looks like. Note that

$$Z_{p,x} = (\sum_{n \le x} |f(n)|^2)^{-1/2} \alpha(p) f(p) \sum_{m \le x/p,\, P(m) < p} \alpha(m) f(m)$$

so

$$\mathbb{E}[|Z_{p,x}|^2 \mid \mathcal{F}_{p^-}] = (\sum_{n \le x} |f(n)|^2)^{-1} |f(p)|^2 |\sum_{m \le x/p,\, P(m) < p} \alpha(m) f(m)|^2.$$

This shows

$$BP_x = (\sum_{n \le x} |f(n)|^2)^{-1} \sum_{p \le x} |f(p)|^2 |\sum_{m \le x/p,\, P(m) < p} \alpha(m) f(m)|^2.$$

We introduce

$$s_{t,x} := t^{-1/2} \sum_{n \le t,\, P(n) < x} \alpha(n) f(n)$$

so that

$$BP_x = \frac{x}{\sum_{n \le x} |f(n)|^2} \sum_{p \le x} \frac{|f(p)|^2}{p} |s_{x/p,p}|^2.$$

It is useful to massage BP_x a little: since $F(x) = \sum_{p \le x} |f(p)|^2/p \sim \theta \log\log x$ then, informally,

$$BP_x \approx \frac{x}{\sum_{n \le x} |f(n)|^2} \int_{2^-}^{x} |s_{x/t,t}|^2 dF(t) \sim \theta \frac{x}{\sum_{n \le x} |f(n)|^2} \int_{2^-}^{x} \frac{|s_{x/t,t}|^2}{t \log t} dt. \quad (1)$$

There are now (at least) two rather different approaches relating this to the measure m_∞. Both approaches employ the following form of Plancherel's identity:

$$\int_{\mathbb{R}} |\sum_{n\le t} g(n)|^2 t^{-2-2r} dt = \frac{1}{2\pi}\int_{\mathbb{R}} |\sum_n g(n) n^{-1/2-r-it}|^2 |1/2+r+it|^{-2} dt \qquad (2)$$

holds for $r>0$, a.s. Proof: Through Perron.
First argument. We introduce

$$U_x := (\sum_{n\le x} |f(n)|^2/n)^{-1} \int_1^x |s_{t,\infty}|^2 \frac{dt}{t}.$$

We combine two lemmas:

$$U_x \xrightarrow{p} \frac{1}{2\pi}\int_{\mathbb{R}} |1/2+it|^{-2} m_\infty(dt) \qquad (3)$$

and

$$\mathbb{E}|U_x - BP_x|^2 \to 0 \qquad (4)$$

The statement (4) is proved by a straightforward technical computation, but isn't deep. The statement (3) is also not too deep (given the construction(s) of m_∞). Perhaps the difficult part is *guessing* that $BP_x \xrightarrow{p} \frac{1}{2\pi}\int_{\mathbb{R}} |1/2 + it|^{-2} m_\infty(dt)$, which is motivated by Plancherel.

To prove (4) one expands the square and computes three expectation using $\mathbb{E}\alpha(n)\overline{\alpha(m)} = \delta_{n,m}$. This leads to summing solutions to $ab=cd$ with certain weights and constraints. Here it is useful to mention some facts:

- Solutions to $ab=cd$ are given by $a=n_1n_2$, $b=n_3n_4$, $c=n_1n_3$, $d=n_2n_4$. Proof: let $g_1=(a,c)$ $g_2=(b,d)$ and define a',b',c',d' accordingly to get $a'b'=c'd'$ with $(a',c')=1$ and $(b',d')=1$. Then $a'=d'$ and $b'=c'$ forcing $a=g_1a$, $b=g_2b'$, $c=g_1b'$, $d=g_2a'$.
- De Bruijn–van Lint: $\sum_{n\le x,\, P(n)\le y} |f(n)|^2 \sim \sum_{n\le x} |f(n)|^2 \cdot \rho_\theta(\frac{\log x}{\log y})$ holds in the regime $\log y \asymp \log x$.
- Wirsing: $\sum_{n\le x} |f(n)|^2 \sim C_f x (\log x)^{\theta-1}$ if $|f(p)|^2 \sim \theta$ on average.

Let us expand on (3). Notation:

$$m_x(dt) = |A(1/2+1/(2\log x)+it)|^2 dt / \mathbb{E}|A(1/2+1/(2\log x)+it)|^2$$

where $A(s)=\sum_n f(n)/n^s$. The measure $m_\infty(dt)$ is constructed by taking $\lim_x m_x(dt)$. Note that

$$\mathbb{E}|A(1/2+r/2+it)|^2 = \mathbb{E}|\sum_n \alpha(n)f(n)/n^{1/2+r/2+it}|^2 = \sum_n |f(n)|^2/n^{1+r}$$

is independent of t. So, for $r=1/\log x$, Plancherel implies that

$$\begin{aligned}\int_{\mathbb{R}} |s_{t,\infty}|^2 t^{-1-1/\log x} dt &= \frac{1}{2\pi}\int_{\mathbb{R}} |A(1/2+1/(2\log x)+it)|^2 |1/2+1/2(1/\log x)+it|^{-2} dt \\ &= \frac{1}{2\pi}(\sum_n |f(n)|^2/n^{1+1/\log x}) \int_{\mathbb{R}} |1/2+1/(2\log x)+it|^{-2} m_x(dt).\end{aligned}$$

From Wirsing, $\sum_n |f(n)|^2/n^{1+1/\log x} \sim C_f\Gamma(\theta)(\log x)^\theta$. From consequences of Plancherel and Wirsing,

$$(C_f\Gamma(\theta)(\log x)^\theta)^{-1}\int_{\mathbb{R}} |s_{t,\infty}|^2 t^{-1-1/\log x} dt \xrightarrow{p} \frac{1}{2\pi}\int_{\mathbb{R}} |1/2+it|^{-2} m_\infty(dt).$$

(Here we relied on $m_x(I) \xrightarrow{p} m_\infty(I)$.) By Tauberian theorem (basically, approximating $\mathbf{1}_{t\le x}$ by $P(t^{-1/\log x})$), this implies that

$$(C_f\Gamma(\theta)(\log x)^\theta)^{-1}\int_1^x |s_{t,\infty}|^2 \frac{dt}{t} \xrightarrow{p} \frac{1}{\Gamma(1+\theta)}\frac{1}{2\pi}\int_{\mathbb{R}} |1/2+it|^{-2} m_\infty(dt)$$

as needed. (In U_x we divide by $\sum_{n\le x} |f(n)|^2/n$, which is asymptotic to $C_f(\log x)^\theta/\theta$ by Wirsing.)

Second argument. Some notation:

$$m_{y,x}(dt) = |A_y(1/2+it+1/(2\log x))|^2 dt/\mathbb{E}|A_y(1/2+it+1/(2\log x))|^2$$

where $A_y(s) = \sum_{P(n)<y} f(n)/n^s$. Recall (1). From Plancherel's identity, we may obtain the limit of a somewhat similar expression, namely

$$x\Big(\sum_{n\le x}|f(n)|^2\Big)^{-1}\int_0^\infty q(t^{1/\log x})\frac{|s_{x/t,x^a}|^2 dt}{t\log x} \tag{5}$$

for any fixed $a>0$ and any 'nice' function q. Details: Plancherel implies that

$$\begin{aligned}\int_{\mathbb{R}}|s_{t,y}|^2 t^{-1-\frac{r}{\log y}}\,dt &= \frac{1}{2\pi}\int_{\mathbb{R}}\Big|A_y\Big(\frac{1}{2}+\frac{r}{2\log y}+it\Big)\Big|^2\Big|\frac{1}{2}+\frac{r}{2\log y}+it\Big|^{-2}dt\\ &= \mathbb{E}\Big|A_y\Big(\frac{1}{2}+\frac{r}{2\log y}\Big)\Big|^2\cdot\frac{1}{2\pi}\int_{\mathbb{R}}\Big|\frac{1}{2}+\frac{r}{2\log y}+it\Big|^{-2}m_{y,y^{1/r}}(dt)\end{aligned}$$

holds for $r>0$ so

$$\Bigg(\sum_{P(n)<y}\frac{|f(n)|^2}{n^{1+r/\log y}}\Bigg)^{-1}\int_0^\infty\frac{|s_{t,y}|^2 dt}{t^{1+r/\log y}}\xrightarrow[y\to\infty]{p} V;$$

now substitute $t=x/t$ and $y=x^a$. This gives (5) with $q(z)=z^{r/a}$. Here we rely on $m_{x,y}(I)\xrightarrow{p} m_\infty(I)$ if $x,y\to\infty$ together. The main difference between (1) and (5), however, is that in (5) the 'smoothness parameter' in $s_{x/t,x^a}$, namely x^a, does not depend on the integration variable t, while in (1) the smoothness parameter in $s_{x/t,t}$ is t, the integration variable itself.

To circumvent this issue, we modify S_x. We divide the primes in $[2,x]$ into finitely many disjoint intervals $(I_k)_k$, and if $n\le x$ has $P(n)\in I_k$, we 'keep' this n in the modified version of S_x only if $P(n/P(n))$ (the second largest prime factor of n) is smaller than $\min I_k$. In this way, the new S_x is

$$S'_x = \sum_{p\le x} Z'_{p,x}$$

where, if $p\in I_k$, then

$$Z'_{p,x} := \Big(\sum_{n\le x}|f(n)|^2\Big)^{-1/2}\sum_{n\le x,\,P(n)=p,\,P(n/P(n))<\min I_k}\alpha(n)f(n)$$

and the new bracket process takes the shape

$$BP'_x \approx x\Big(\sum_{n\le x}|f(n)|^2\Big)^{-1}\sum_k\sum_{p\in I_k}\frac{|f(p)|^2}{p}|s_{x/p,\min I_k}|^2 \approx x\Big(\sum_{n\le x}|f(n)|^2\Big)^{-1}\sum_k\int_2^x\frac{|s_{x/t,\min I_k}|^2 dt}{t\log t}.$$

For fixed k, the smoothness parameter is now fixed within the integral, namely it is $\min I_k$. This allows us to handle the kth integral using (5). One has to justify working with this modified S_x. The idea is that the second moment of the discarded terms is

$$\sum_k\frac{\sum_{n\le x:\,P(n),P(n/P(n))\in I_k}|f(n)|^2}{\sum_{n\le x}|f(n)|^2}\ll\sum_k\sum_{p,q\in I_k}\frac{1}{pq}\ll\sum_k(\log(\log\max I_k/\log\min I_k))^2.$$

We discard also $P(n)\le x^\varepsilon$ (we lose $O(\varepsilon)$) and then take $I_k=[x^{\varepsilon+\delta k},x^{\varepsilon+\delta(k+1)}]$, so that the summand is $\ll \log^2((\varepsilon+\delta(k+1))/(\varepsilon+\delta k))\ll\min\{\delta/\varepsilon,1/k\}^2$ and this is good enough; the total loss is $O_\varepsilon(\delta)+O(\varepsilon)$ and this is manageable if we take δ to 0 and then ε to 0.