

# SMS 2026: Universal Statistics in Number Theory Probability Working Sessions

Louis-Pierre Arguin & Christian Webb

May 5-7 2026

## Abstract

This is the companion worksheet to the lecture series *Introduction to Multiplicative Chaos* by Christian Webb and *Extrema of log-correlated processes* by Louis-Pierre Arguin during the Séminaire de Mathématiques Supérieures 2026 *Universal Statistics in Number Theory* at the Centre de Recherches Mathématiques, Montréal.

## Contents

<b>1</b>	<b>Computations related to multiplicative chaos</b>	<b>2</b>
1.1	Logarithmic correlations in the Steinhaus model . . . . .	2
1.2	Euler product estimates . . . . .	5
1.3	A Gaussian calculation . . . . .	7
1.4	Estimates needed for multiplicative chaos . . . . .	8
1.5	Hints . . . . .	10
<b>2</b>	<b>Estimates of the Steinhaus Model of <math>\zeta</math></b>	<b>11</b>
2.1	Some Useful Estimates . . . . .	11
2.2	Moments of the Steinhaus model . . . . .	13
<b>3</b>	<b>Ballot Theorems</b>	<b>15</b>
3.1	Reflection Principle . . . . .	15
3.2	Ballot theorems for Gaussian random walks. . . . .	17
<b>4</b>	<b>Gaussian Approximation and Harmonic Analysis</b>	<b>19</b>
4.1	Berry-Esseen Estimates . . . . .	19
4.2	Discretization . . . . .	22

# 1 Computations related to multiplicative chaos

This section covers the exercises for the course about multiplicative chaos. There is some overlap with the exercises about the extrema, and a bit fewer problems. There are also some hints at the end.

## 1.1 Logarithmic correlations in the Steinhaus model

The first exercise is about digging into the details of why we see logarithmic correlations in the Steinhaus model on the critical line. This is a bit of work and don't worry if you don't get all the way to the end.

**1.** Let  $\mathbf{P}$  denote the set of prime numbers and for  $P \in \mathbf{P}$ ,  $\mathbf{P}_P = \{p \in \mathbf{P} : p \leq P\}$ . Let  $(f(p))_{p \in \mathbf{P}}$  be independent Steinhaus random variables and for  $t \in \mathbb{R}$

$$X_P(t) = \log \left( \prod_{p \in \mathbf{P}_P} \frac{1}{|1 - f(p)p^{-\frac{1}{2}-it}|} \right).$$

Show that  $\mathbb{E}[X_P(t)] = 0$  and uniformly in  $t, t' \in K$ , where  $K \in \mathbb{R}$  is compact,

$$\mathbb{E}[X_P(t)X_P(t')] = \frac{1}{2} \log(\min(|t - t'|^{-1}, \log P)) + O(1)$$

as  $P \rightarrow \infty$ .

*Proof.* Recall that for a complex number  $z \neq 0$ , we have (for any branch of the logarithm),  $\log |z| = \operatorname{Re}(\log z)$ , and if  $|z| < 1$ , we have for the principal branch of the logarithm  $-\log(1 - z) = \sum_{n=1}^{\infty} \frac{1}{n} z^n$ . Thus (as  $|f(p)p^{-\frac{1}{2}-it}| \leq 2^{-1/2} < 1$ )

$$X_P(t) = \log \left( \prod_{p \in \mathbf{P}_P} \frac{1}{|1 - f(p)p^{-\frac{1}{2}-it}|} \right) = \frac{1}{2} \sum_{p \in \mathbf{P}_P} \sum_{n=1}^{\infty} \frac{1}{n} p^{-\frac{n}{2}} (f(p)^n p^{-int} + f(p)^{-n} p^{int}).$$

As this series converges uniformly in everything, we can use Fubini to compute

$$\mathbb{E}[X_P(t)] = \frac{1}{2} \sum_{p \in \mathbf{P}_P} \sum_{n=1}^{\infty} \frac{1}{n} p^{-\frac{n}{2}} (\mathbb{E}[f(p)^n] p^{-int} + \mathbb{E}[f(p)^{-n}] p^{int}).$$

Since for any integer  $k$  we have  $\mathbf{E}[f(p)^k] = \delta_{k,0}$ , we see that  $\mathbb{E}[f(p)^n] = \mathbb{E}[f(p)^{-n}] = 0$  for  $n \geq 1$  so

$$\mathbb{E}[X_P(t)] = 0.$$

This concludes the proof of the first claim.

The second one is lengthier. With start with a similar Fubini argument to argue that

$$\begin{aligned} \mathbb{E}[X_P(t)X_P(t')] &= \frac{1}{4} \sum_{p,q \in \mathbf{P}_P} \sum_{n,m=1}^{\infty} \frac{1}{nm} p^{-\frac{n}{2}} q^{-\frac{m}{2}} \mathbb{E} \left[ f(p)^n f(q)^m p^{-int} q^{-imt'} + f(p)^n f(q)^{-m} p^{-int} q^{imt'} \right. \\ &\quad \left. + f(p)^{-n} f(q)^m p^{int} q^{-imt'} + f(p)^{-n} f(q)^{-m} p^{int} q^{imt'} \right]. \end{aligned}$$

Recalling that for any integers  $k, k'$  and distinct primes  $p$  and  $q$ , we have  $\mathbb{E}[f(p)^k f(q)^{k'}] = \delta_{k,0} \delta_{k',0}$ , we see that  $\mathbb{E}[f(p)^n f(q)^m] = \mathbb{E}[f(p)^{-n} f(q)^{-m}] = 0$  and  $\mathbb{E}[f(p)^n f(q)^{-m}] = \mathbb{E}[f(p)^{-n} f(q)^m] = \delta_{p,q} \delta_{m,n}$ . This means that

$$\mathbb{E}[X_P(t)X_P(t')] = \frac{1}{4} \sum_{p \in \mathbf{P}_P} \sum_{n=1}^{\infty} \frac{1}{n^2} p^{-n} (p^{-in(t-t')} + p^{in(t-t')}).$$

Note that for  $n \geq 2$  part the sum we have

$$\left| \frac{1}{4} \sum_{p \in \mathbf{P}_P} \sum_{n=2}^{\infty} \frac{1}{n^2} p^{-n} (p^{-in(t-t')} + p^{in(t-t')}) \right| \leq \frac{1}{4} \sum_{p \in \mathbf{P}_P} p^{-2} \sum_{n=2}^{\infty} \frac{1}{n^2} < \infty$$

so this produces only something  $O(1)$  to our covariance and it is sufficient for us to focus on the  $n = 1$  sum.

This we write in terms of the prime counting function  $\pi(n)$  and sum by parts

$$\begin{aligned} \frac{1}{4} \sum_{p \in \mathbf{P}_P} p^{-1} (p^{-i(t-t')} + p^{i(t-t')}) &= \frac{1}{2} \sum_{n=2}^P \frac{\cos((t-t') \log n)}{n} (\pi(n) - \pi(n-1)) \\ &= \frac{1}{2} \sum_{n=2}^{P-1} \left( \frac{\cos((t-t') \log n)}{n} - \frac{\cos((t-t') \log(n+1))}{n+1} \right) \pi(n) \\ &\quad + \frac{1}{2} \frac{\cos((t-t') \log P)}{P} \pi(P). \end{aligned}$$

We now recall the prime number theorem (in a quantitative form) which states that as  $n \rightarrow \infty$ ,

$$\pi(n) = \frac{n}{\log n} + O\left(\frac{n}{(\log n)^2}\right).$$

We see immediately that the  $\pi(P)$ -term above is  $o(1)$ . We also note since

$$\partial_x \frac{\cos((t-t') \log x)}{x} = - \frac{(t-t') \sin((t-t') \log x) + \cos((t-t') \log x)}{x^2}$$

and

$$\left| \partial_x^2 \frac{\cos((t-t') \log x)}{x} \right| \leq \frac{4(1 + |t-t'|)^2}{x^3},$$

there exists a constant  $C = C_K$  (depending only on the compact set  $K$  that  $t, t'$  are in) such that

$$\begin{aligned} &\left| \frac{\cos((t-t') \log n)}{n} - \frac{\cos((t-t') \log(n+1))}{n+1} - \frac{(t-t') \sin((t-t') \log n) + \cos((t-t') \log n)}{n^2} \right| \\ &\leq C_K \frac{1}{n^3}. \end{aligned}$$

This means that

$$\begin{aligned} & \frac{1}{2} \sum_{n=2}^{P-1} \left( \frac{\cos((t-t') \log n)}{n} - \frac{\cos((t-t') \log(n+1))}{n+1} \right) \pi(n) \\ &= \frac{1}{2} \sum_{n=10}^{P-1} \frac{\cos((t-t') \log n) + (t-t') \sin((t-t') \log n)}{n \log n} + O(1) + O\left(\sum_{n=10}^{P-1} \frac{1}{n(\log n)^2}\right), \end{aligned}$$

where the implied constants may depend on  $K$ . The last term can be compared to the integral

$$\int_5^\infty \frac{1}{x(\log x)^2} dx = \int_{\log 5}^\infty \frac{1}{y^2} dy < \infty$$

(where we performed the change of variables  $x = e^y$ ) so

$$\begin{aligned} & \frac{1}{2} \sum_{n=2}^{P-1} \left( \frac{\cos((t-t') \log n)}{n} - \frac{\cos((t-t') \log(n+1))}{n+1} \right) \pi(n) \\ &= \frac{1}{2} \sum_{n=10}^{P-1} \frac{\cos((t-t') \log n) + (t-t') \sin((t-t') \log n)}{n \log n} + O(1). \end{aligned}$$

We can perform a similar integral estimate for our first sum. By another routine derivative estimate as before

$$\begin{aligned} & \frac{\cos((t-t') \log n) + (t-t') \sin((t-t') \log n)}{n \log n} \\ &= \int_n^{n+1} \frac{\cos((t-t') \log x) + (t-t') \sin((t-t') \log x)}{x \log x} dx + O\left(\int_n^{n+1} \frac{1}{x^2 \log x} dx\right) \\ &= \int_n^{n+1} \frac{\cos((t-t') \log x) + (t-t') \sin((t-t') \log x)}{x \log x} dx + O(n^{-2}) \end{aligned}$$

so

$$\begin{aligned} & \frac{1}{2} \sum_{n=10}^{P-1} \frac{\cos((t-t') \log n) + (t-t') \sin((t-t') \log n)}{n \log n} \\ &= \frac{1}{2} \int_{10}^P \frac{\cos((t-t') \log x) + (t-t') \sin((t-t') \log x)}{x \log x} dx + O(1) \end{aligned}$$

(where again the implied constant can depend on  $K$ , but not  $P$  or  $t, t'$  otherwise). Writing

again  $x = e^y$ , we can write our integral as

$$\begin{aligned}
& \frac{1}{2} \int_{10}^P \frac{\cos((t-t') \log x) + (t-t') \sin((t-t') \log x)}{x \log x} dx \\
&= \frac{1}{2} \int_{\log 10}^{\log P} \frac{\cos((t-t')y) + (t-t') \sin((t-t')y)}{y} dy \\
&= \frac{1}{2} \int_1^{\log P} \frac{\cos((t-t')y) + (t-t') \sin((t-t')y)}{y} dy + O(1) \\
&= \frac{1}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{\cos(y)}{y} dy + \frac{|t-t'|}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{\sin(y)}{y} dy.
\end{aligned}$$

If  $|t-t'| \log P \leq 1$ , then we write in the first integral  $\frac{\cos(y)}{y} = \frac{1}{y} + O(y)$  and in the second integral  $\frac{\sin(y)}{y} = O(1)$ , to find

$$\begin{aligned}
\frac{1}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{\cos(y)}{y} dy + \frac{|t-t'|}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{\sin(y)}{y} dy &= \frac{1}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{dy}{y} + O\left(\int_0^1 (1+y) dy\right) \\
&= \frac{1}{2} \log \log P + O(1),
\end{aligned}$$

where once again, the implied constant depends only on  $K$ . This concludes the proof of the claim in the case  $|t-t'| \log P \leq 1$ . If on the other hand  $|t-t'| \log P \geq 1$ , we have

$$\begin{aligned}
& \frac{1}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{\cos(y)}{y} dy + \frac{|t-t'|}{2} \int_{|t-t'|}^{|t-t'| \log P} \frac{\sin(y)}{y} dy \\
&= \frac{1}{2} \int_{|t-t'|}^1 \frac{dy}{y} + O\left(\int_0^1 (1+y) dy\right) + \frac{1}{2} \int_1^{|t-t'| \log P} \frac{\cos(y)}{y} dy + \frac{|t-t'|}{2} \int_1^{|t-t'| \log P} \frac{\sin(y)}{y} dy.
\end{aligned}$$

The first integral is simply  $-\frac{1}{2} \log |t-t'|$ , the second quantity is  $O(1)$ , while the last two integrals are perhaps slightly less trivial, they are known to be bounded in  $P$  and  $t, t'$  (see e.g. [https://en.wikipedia.org/wiki/Trigonometric\\_integral](https://en.wikipedia.org/wiki/Trigonometric_integral)). To summarize, we have argued that

$$\mathbb{E}[X_P(t)X_P(t')] = \frac{1}{2} \log(\min(|t-t'|^{-1}, \log P)) + O(1)$$

where the implied constant depends only on the compact set  $K$ . □

## 1.2 Euler product estimates

We now focus on perhaps the main estimate for constructing multiplicative chaos measures from the Steinhaus model.

2. Using similar notation as in the previous problem, let  $k \geq 1$ ,  $\alpha_1, \dots, \alpha_k \in \mathbb{R}$  be fixed (independent of  $P$ ) and  $100(1 + (\sum_{j=1}^k |\alpha_j|)^2) \leq Q < P$ . Show that uniformly in  $t_1, \dots, t_k \in \mathbb{R}$

$$\begin{aligned} & \mathbb{E} \left[ e^{\sum_{j=1}^k 2\alpha_j (X_P(t_j) - X_Q(t_j))} \right] \\ &= \mathbb{E} \left[ \prod_{j=1}^k \prod_{Q < p \leq P} \frac{1}{|1 - f(p)p^{-\frac{1}{2} - it_j}|^{2\alpha_j}} \right] \\ &= \exp \left( \sum_{j=1}^k \alpha_j^2 \sum_{Q < p \leq P} \frac{1}{p} + 2 \sum_{1 \leq l < j \leq k} \alpha_j \alpha_l \sum_{Q < p \leq P} \frac{\cos((t_l - t_j) \log p)}{p} + O(1) \right). \end{aligned}$$

*Proof.* We start by using independence to write this as

$$\mathbb{E} \left[ \prod_{j=1}^k \prod_{Q < p \leq P} \frac{1}{|1 - f(p)p^{-\frac{1}{2} - it_j}|^{2\alpha_j}} \right] = \prod_{Q < p \leq P} \mathbb{E} \left[ e^{-\sum_{j=1}^k 2\alpha_j \log |1 - f(p)p^{-\frac{1}{2} - it_j}|} \right].$$

We then use the same old series expansion of the logarithm:

$$-2 \log |1 - f(p)p^{-\frac{1}{2} - it_j}| = \sum_{n=1}^{\infty} \frac{1}{n} p^{-\frac{n}{2}} (f(p)^n p^{-int_j} + f(p)^{-n} p^{int_j}).$$

We cannot now directly argue that the  $n \geq 2$  part can be absorbed into the  $O(1)$ , but we can do this for  $n \geq 3$ . The reason for this being that

$$\begin{aligned} \sum_{p \in \mathbf{P}} \sum_{n=3}^{\infty} \frac{1}{n} p^{-\frac{n}{2}} |f(p)^n p^{-int_j} + f(p)^{-n} p^{int_j}| &\leq 2 \sum_{p \in \mathbf{P}} p^{-\frac{4}{3}} \sum_{n=3}^{\infty} \frac{1}{n} p^{-\frac{n}{20}} \\ &\leq \sum_{k=3}^{\infty} k^{-4/3} \sum_{j=3}^{\infty} 2^{-\frac{j}{20}} \\ &< \infty \end{aligned}$$

uniformly in  $f(p)$  and  $t_j$ . Here we used the fact that  $p^{-\frac{4}{3} - \frac{n}{20}} \geq p^{-\frac{n}{2}}$  for  $n \geq 3$  (since  $\frac{4}{3} + \frac{n}{20} \leq \frac{n}{2}$  for  $n \geq 3$ ). Thus we can absorb (deterministically), the sum over  $n \geq 3$  into the  $O(1)$  term in the claim. Thus our task is to control

$$\prod_{Q < p \leq P} \mathbb{E} \left[ e^{\sum_{j=1}^k \alpha_j (p^{-\frac{1}{2}} (f(p)p^{-it_j} + f(p)^{-1} p^{it_j}) + \frac{1}{2} p^{-1} (f(p)^2 p^{-2it_j} + f(p)^{-2} p^{2it_j}))} \right]$$

We now expand the exponential to third order ( $e^x = 1 + x + \frac{1}{2}x^2 + O(x^3)$ ) to find

$$\begin{aligned} & e^{\sum_{j=1}^k \alpha_j (p^{-\frac{1}{2}} (f(p)p^{-it_j} + f(p)^{-1} p^{it_j}) + \frac{1}{2} p^{-1} (f(p)^2 p^{-2it_j} + f(p)^{-2} p^{2it_j}))} \\ &= 1 + \sum_{j=1}^k \alpha_j \left( p^{-\frac{1}{2}} (f(p)p^{-it_j} + f(p)^{-1} p^{it_j}) + \frac{1}{2} p^{-1} (f(p)^2 p^{-2it_j} + f(p)^{-2} p^{2it_j}) \right) \\ &+ \frac{1}{2} \sum_{l,j=1}^k \alpha_l \alpha_j p^{-1} (f(p)p^{-it_l} + f(p)^{-1} p^{it_l}) (f(p)p^{-it_j} + f(p)^{-1} p^{it_j}) + O(p^{-3}), \end{aligned}$$

where the implied constant is uniform in  $f(p)$  and  $t_1, \dots, t_k$ . Taking expectations and recalling that  $\mathbf{E}[f(p)^n] = \delta_{n,0}$  for an integer  $n$ , we see that

$$\begin{aligned}
& \mathbb{E} \left[ e^{\sum_{j=1}^k \alpha_j (p^{-\frac{1}{2}}(f(p)p^{-it_j} + f(p)^{-1}p^{it_j}) + \frac{1}{2}p^{-1}(f(p)^2p^{-2it_j} + f(p)^{-2}p^{-2it_j}))} \right] \\
&= 1 + \sum_{j=1}^k \frac{\alpha_j^2}{p} + 2 \sum_{1 \leq j < l \leq k} \alpha_j \alpha_l \frac{\cos((t_j - t_l) \log p)}{p} + O(p^{-3/2}) \\
&= e^{\log(1 + \sum_{j=1}^k \frac{\alpha_j^2}{p} + 2 \sum_{1 \leq j < l \leq k} \alpha_j \alpha_l \frac{\cos((t_j - t_l) \log p)}{p})} + O(p^{-3/2}) \\
&= e^{\sum_{j=1}^k \frac{\alpha_j^2}{p} + 2 \sum_{1 \leq j < l \leq k} \alpha_j \alpha_l \frac{\cos((t_j - t_l) \log p)}{p}} + O(p^{-3/2}).
\end{aligned}$$

Taking the product over  $p$  and noting that  $\sum_{p \in \mathbf{P}} p^{-3/2} < \infty$ , we conclude that

$$\begin{aligned}
& \mathbb{E} \left[ \prod_{j=1}^k \prod_{Q < p \leq P} \frac{1}{|1 - f(p)p^{-\frac{1}{2} - it_j}|^{2\alpha_j}} \right] \\
&= \exp \left( \sum_{j=1}^k \alpha_j^2 \sum_{Q < p \leq P} \frac{1}{p} + 2 \sum_{1 \leq l < j \leq k} \alpha_j \alpha_l \sum_{Q < p \leq P} \frac{\cos((t_l - t_j) \log p)}{p} + O(1) \right)
\end{aligned}$$

as claimed.  $\square$

### 1.3 A Gaussian calculation

We now argue that the last exercise is essentially saying that  $X_P$  behaves like a Gaussian object.

**3.** Let  $(U_p, V_p)_{p \in \mathbf{P}}$  be independent standard Gaussian random variables. Define for  $P \in \mathbf{P}$

$$Y_P(t) = \frac{1}{\sqrt{2}} \sum_{p \in \mathbf{P}_P} p^{-1/2} (U_p \cos(t \log p) + V_p \sin(t \log p)).$$

Compute for  $k \geq 1$ ,  $\alpha_1, \dots, \alpha_k \in \mathbb{R}$  and  $2 \leq Q < P$  (with  $Q \in \mathbf{P}$ ), and  $t_1, \dots, t_k \in \mathbb{R}$

$$\mathbb{E} \left[ e^{\sum_{j=1}^k 2\alpha_j (Y_P(t_j) - Y_Q(t_j))} \right].$$

*Proof.* We begin by writing

$$\begin{aligned}
\sum_{j=1}^k 2\alpha_j (Y_P(t_j) - Y_Q(t_j)) &= \sum_{Q < p \leq P} U_p \left( \frac{1}{\sqrt{2}} p^{-1/2} \sum_{j=1}^k 2\alpha_j \cos(t_j \log p) \right) \\
&\quad + \sum_{Q < p \leq P} V_p \left( \frac{1}{\sqrt{2}} p^{-1/2} \sum_{j=1}^k 2\alpha_j \sin(t_j \log p) \right).
\end{aligned}$$

By independence, we thus have

$$\begin{aligned} & \mathbb{E} \left[ e^{\sum_{j=1}^k 2\alpha_j (Y_P(t_j) - Y_Q(t_j))} \right] \\ &= \prod_{Q < p \leq P} \mathbb{E} \left[ e^{U_p \left( \frac{1}{\sqrt{2}} p^{-1/2} \sum_{j=1}^k 2\alpha_j \cos(t_j \log p) \right)} \right] \mathbb{E} \left[ e^{V_p \left( \frac{1}{\sqrt{2}} p^{-1/2} \sum_{j=1}^k 2\alpha_j \sin(t_j \log p) \right)} \right]. \end{aligned}$$

We now recall that for a standard Gaussian random variable  $Z$  (such as  $U_p$  or  $V_p$ ) and  $\alpha \in \mathbb{R}$ , we have

$$\begin{aligned} \mathbb{E}[e^{\alpha Z}] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\alpha x} e^{-\frac{x^2}{2}} dx \\ &= e^{\frac{\alpha^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-\alpha)^2}{2}} dx \\ &= e^{\frac{\alpha^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx \\ &= e^{\frac{\alpha^2}{2}}. \end{aligned}$$

Thus

$$\begin{aligned} & \mathbb{E} \left[ e^{\sum_{j=1}^k 2\alpha_j (Y_P(t_j) - Y_Q(t_j))} \right] \\ &= \prod_{Q < p \leq P} e^{\frac{1}{p} (\sum_{j=1}^k \alpha_j \cos(t_j \log p))^2 + \frac{1}{p} (\sum_{j=1}^k \alpha_j \sin(t_j \log p))^2} \\ &= \prod_{Q < p \leq P} e^{\frac{1}{p} \sum_{j=1}^k \alpha_j^2 (\cos^2(t_j \log p) + \sin^2(t_j \log p)) + \frac{2}{p} \sum_{1 \leq j < l \leq k} \alpha_j \alpha_l (\cos(t_j \log p) \cos(t_l \log p) + \sin(t_j \log p) \sin(t_l \log p))} \\ &= \prod_{Q < p \leq P} e^{\frac{1}{p} \sum_{j=1}^k \alpha_j^2 + \frac{2}{p} \sum_{1 \leq j < l \leq k} \alpha_j \alpha_l \cos((t_j - t_l) \log p)} \\ &= \exp \left( \sum_{j=1}^k \alpha_j^2 \sum_{Q < p \leq P} \frac{1}{p} + 2 \sum_{1 \leq l < j \leq k} \alpha_j \alpha_l \sum_{Q < p \leq P} \frac{\cos((t_l - t_j) \log p)}{p} \right), \end{aligned}$$

which is exactly what we found in the last exercise apart from the  $O(1)$  term.  $\square$

## 1.4 Estimates needed for multiplicative chaos

Finally we combine some of the work we have done to derive some important estimates in our construction of subcritical multiplicative chaos measures.

4.

1. Prove that for fixed  $\gamma > 0$

$$\mathbb{E}[e^{\gamma X_P(t)}] = e^{\frac{\gamma^2}{4} \log \log P + O(1)}$$

uniformly in  $t \in \mathbb{R}$ .

2. Prove that for fixed  $\gamma > 0$

$$\frac{\mathbb{E}[e^{\gamma X_P(t) + \gamma X_P(t')}]}{\mathbb{E}[e^{\gamma X_P(t)}]\mathbb{E}[e^{\gamma X_P(t')}]}\leq e^{O(1)}|t - t'|^{-\frac{\gamma^2}{2}}$$

uniformly in  $t, t'$  in a fixed compact set  $K \subset \mathbb{R}$ .

3. Prove that for fixed  $\alpha, \gamma > 0$

$$\frac{\mathbb{E}[e^{\gamma X_P(t) + \alpha X_Q(t)}]}{\mathbb{E}[e^{\gamma X_P(t)}]}\leq e^{O(1) + \frac{\alpha^2}{4} \log \log Q + \frac{\alpha\gamma}{2} \log \log Q}$$

uniformly in  $t \in \mathbb{R}$  and  $10 \leq Q < P$ .

*Proof.* 1. We use Problem 2 with  $k = 1$ ,  $2\alpha_1 = \gamma$  and  $Q \geq 100(1 + \frac{\gamma^2}{4})$  (the closest prime to this quantity). We then have by independence

$$\mathbb{E}[e^{\gamma X_P(t)}] = \mathbb{E}[e^{\gamma X_Q(t)}]\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t))}].$$

Since  $Q$  is fixed (independent of  $P$ ), the prefactor is bounded from above and away from zero in  $P$  and contributes  $e^{O(1)}$ . For the second expectation, Problem 2 yields

$$\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t))}] = e^{\frac{\gamma^2}{4} \sum_{Q < p \leq P} \frac{1}{p} + O(1)} = e^{\frac{\gamma^2}{4} \sum_{2 \leq p \leq P} \frac{1}{p} + O(1)}$$

again since  $Q$  is fixed. By Merten's third theorem,  $\sum_{2 \leq p \leq P} \frac{1}{p} = \log \log P + O(1)$ , so we find that

$$\mathbb{E}[e^{\gamma X_P(t)}] = e^{\frac{\gamma^2}{4} \log \log P + O(1)}$$

as claimed.

2. We again use Problem 2, but now with  $k = 2$ ,  $2\alpha_1 = 2\alpha_2 = \gamma$  and  $Q \geq 100(1 + \gamma^2)$  (the closest prime to this quantity). We then find by independence

$$\frac{\mathbb{E}[e^{\gamma X_P(t) + \gamma X_P(t')}]}{\mathbb{E}[e^{\gamma X_P(t)}]\mathbb{E}[e^{\gamma X_P(t')}]}\frac{\mathbb{E}[e^{\gamma X_Q(t) + \gamma X_Q(t')}]}{\mathbb{E}[e^{\gamma X_Q(t)}]\mathbb{E}[e^{\gamma X_Q(t')}]}\frac{\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t)) + \gamma(X_P(t') - X_Q(t'))}]}{\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t))}]\mathbb{E}[e^{\gamma(X_P(t') - X_Q(t'))}]}.$$

The first ratio is bounded in  $P$  and we can bound it by say  $e^{O(1)}$  (with the required uniformity). For the second ratio, Problem 2 yields

$$\begin{aligned} & \frac{\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t)) + \gamma(X_P(t') - X_Q(t'))}]}{\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t))}]\mathbb{E}[e^{\gamma(X_P(t') - X_Q(t'))}]} \\ &= e^{\frac{\gamma^2}{4} \sum_{Q < p \leq P} \frac{1}{p} + \frac{\gamma^2}{4} \sum_{Q < p \leq P} \frac{1}{p} + \frac{\gamma^2}{2} \sum_{Q < p \leq P} \frac{\cos((t-t') \log p)}{p} + O(1) - \frac{\gamma^2}{4} \sum_{Q < p \leq P} \frac{1}{p} - \frac{\gamma^2}{4} \sum_{Q < p \leq P} \frac{1}{p} + O(1)} \\ &= e^{\frac{\gamma^2}{2} \sum_{Q < p \leq P} \frac{\cos((t-t') \log p)}{p} + O(1)} \\ &= e^{\frac{\gamma^2}{2} \sum_{2 \leq p \leq P} \frac{\cos((t-t') \log p)}{p} + O(1)} \end{aligned}$$

This is precisely the sum we encountered in Problem 1 and found that

$$\sum_{2 \leq p \leq P} \frac{\cos((t-t') \log p)}{p} = \log(\min(|t-t'|^{-1}, \log P)) + O(1) \leq \log |t-t'|^{-1} + O(1)$$

so we find the claim.

3. Let us use independence to write

$$\frac{\mathbb{E}[e^{\gamma X_P(t) + \alpha X_Q(t)}]}{\mathbb{E}[e^{\gamma X_P(t)}]} = \frac{\mathbb{E}[e^{\gamma X_Q(t) + \alpha X_Q(t)}]}{\mathbb{E}[e^{\gamma X_Q(t)}]} \frac{\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t))}]}{\mathbb{E}[e^{\gamma(X_P(t) - X_Q(t))}]} = \frac{\mathbb{E}[e^{(\gamma + \alpha) X_Q(t)}]}{\mathbb{E}[e^{\gamma X_Q(t)}]}.$$

The claim now follows directly from part 1 of this exercise. □

## 1.5 Hints

### Beware of typos in the hints

1. The first claim about the expectation being zero should be quite straightforward. Just argue that you can series expand the logarithm (recall that  $\log |z| = \operatorname{Re}(\log z)$  for  $z \in \mathbb{C}$  and for  $z \in \mathbb{C}$  with  $|z| < 1$ ,  $-\log(1-z) = \sum_{n=1}^{\infty} \frac{1}{n} z^n$ ). The second claim is a bit more elaborate. Using the same series expansion, try to argue (e.g. using Fubini) that

$$\mathbb{E}[X_P(t) X_P(t')] = \frac{1}{2} \sum_{n=1}^{\infty} \sum_{p \in \mathbf{P}_P} \frac{1}{n^2} p^{-n} \cos(n(t-t') \log p).$$

Then try to argue that the sum over  $n \geq 2$  produces something  $O(1)$  uniformly in  $P, t, t'$ , so you are left with estimating

$$\frac{1}{2} \sum_{p \in \mathbf{P}_P} \frac{\cos((t-t') \log p)}{p}.$$

You can then write this sum as

$$\frac{1}{2} \sum_{n=2}^P \frac{\cos((t-t') \log n)}{n} (\pi(n) - \pi(n-1)),$$

where  $\pi$  is the prime counting function. Perform summation by parts, use some basic derivative estimates, and use the prime number theorem with the estimate

$$\pi(n) = \frac{n}{\log n} + O\left(\frac{n}{(\log n)^2}\right)$$

to write this as something like

$$\frac{1}{4} \sum_{n=10}^{P-1} \left( \frac{1 + i(t-t')}{n^{1+i(t-t')}} + \frac{1 - i(t-t')}{n^{1-i(t-t')}} \right) \frac{1}{\log n} + O(1).$$

Now argue why at the cost of  $O(1)$ , you can replace the sums by integrals and perform a couple of changes of variables in the integrals to arrive at

$$\frac{1}{2} \int_{|t-t'|}^{|t-t'|\log P} \frac{\cos v}{v} dv + \frac{|t-t'|}{2} \int_{|t-t'|}^{|t-t'|\log P} \frac{\sin v}{v} dv + O(1).$$

There are now some regimes to consider. For  $|t-t'|$  bounded from below, you might want to look up sine- and cosine-integrals to justify the claim ([https://en.wikipedia.org/wiki/Trigonometric\\_integral](https://en.wikipedia.org/wiki/Trigonometric_integral)) and their asymptotic properties to conclude (or perform some explicit estimates if you can come up with them). For  $|t-t'|$  tending to zero but slower than  $|t-t'|\log P \geq 1$ , split the integrals at 1, use the sine- and cosine-integrals to estimate the integrals over  $v \geq 1$ , and write  $\cos(v) = 1 + O(v^2)$  and  $\sin(v) = O(v)$  to estimate the small  $v$  integrals. For  $|t-t'|\log P \leq 1$ , argue as in the last case, but you only need to consider the small  $v$  regime.

2. This is a special case of “Euler product result 1” on page 10 of <https://arxiv.org/pdf/2012.15809>.
3. The main idea is to use independence and the fact that  $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\alpha x} e^{-\frac{x^2}{2}} dx = e^{\frac{\alpha^2}{2}}$ . If you have never seen this identity, try to prove it (hint: complete the square).
4. Here you should choose  $Q$  to be of order one and argue why the first  $Q$  terms do not matter in the product. Also you might want to brush up on Mertens’ third theorem for the first claim. For the second claim, it is not directly the result of the first exercise we are using, but an estimate derived in the solution of the first exercise. For the third part, it may be a good idea to write  $X_P(t) = X_Q(t) + X_P(t) - X_Q(t)$  and use independence.

## 2 Estimates of the Steinhaus Model of $\zeta$

### 2.1 Some Useful Estimates

1. **Weyl’s Criterion.** Consider  $r$  distinct primes  $(p_j, j \leq r)$  and  $r$ -tuples  $\mathbf{k} = (k_1, \dots, k_r)$  and  $\mathbf{l} = (\ell_1, \dots, \ell_r)$  of non-negative integers. Prove that

$$\frac{1}{T} \int_0^T \prod_{j=1}^r (p_j^{-it})^{k_j} \overline{(p_j^{-it})^{\ell_j}} dt = \begin{cases} 1 & \text{if } \mathbf{k} = \mathbf{l}, \\ O(1/T) & \text{if } \mathbf{k} \neq \mathbf{l}. \end{cases}$$

Conclude that  $(p^{-i\tau}, p \text{ primes})$  converges to IID Steinhaus random variables as  $T \rightarrow \infty$  in the sense of finite-dimensional distribution.

*Proof.* The integral reduces to

$$\frac{1}{T} \int_0^T \exp(it \log m/n) dt$$

where  $n = p_1^{k_1} \dots p_r^{k_r}$  and  $m = p_1^{-\ell_1} \dots p_r^{-\ell_r}$ . The integral is 1 if  $m = n$  which is equivalent to  $\mathbf{k} = \mathbf{l}$  by unique factorization. It is  $O(1/T)$  otherwise (for  $r$  fixed), which is 0 in the limit  $T \rightarrow \infty$ . This is the result we would obtain replacing  $p^{-it}$  by independent Steinhaus random variables  $e^{i\theta(p)}$ . This proves convergence of the joint moments of the variables which implies convergence of the finite-dimensional distributions (since the unit circle is compact).  $\square$

2. **Log-Correlations.** The Prime Number Theorem with classical error states that

$$\pi(x) = \sum_{p \leq x} 1 = \int_2^x \frac{1}{\log y} dy + O(xe^{-c\sqrt{\log x}}).$$

Use this and summation by parts to show that for  $\delta > 0$  and  $2 \leq x < y$ , we have

$$\sum_{x < p \leq y} \frac{\cos(\delta \log p)}{p} = \begin{cases} \log \log y - \log \log x + O((\delta \log y)^2) + O(e^{-c\sqrt{\log x}}), & \text{if } \delta \log y < 1, \\ O((\delta \log x)^{-1}) + O((1 + \delta)e^{-c\sqrt{\log x}}), & \text{if } \delta \log x \geq 1. \end{cases} \quad (1)$$

*Proof.* Summation by parts gives

$$\sum_{x \leq p \leq y} f(p) = \int_x^y \frac{f(u)}{\log u} du + \mathcal{E}(y)f(y) - \mathcal{E}(x)f(x) - \int_x^y f'(u)\mathcal{E}(u)du, \quad (2)$$

where  $\mathcal{E}(x) = \pi(x) - \text{li}(x) = O(xe^{-c\sqrt{\log x}})$ . We use this with  $f(u) = \frac{\cos(\delta \log u)}{u}$ . We prove the case  $0 < \delta < 1$ . This gives us

$$\begin{aligned} \sum_{x < p \leq y} \frac{\cos(\delta \log p)}{p} &= \int_x^y \frac{\cos(\delta \log u)}{u \log u} du + O\left(e^{-c\sqrt{\log x}} + \int_x^y \frac{\cos(\delta \log u) + \delta \sin(\delta \log u)}{u} e^{-c\sqrt{\log u}} du\right) \\ &= \int_{\log x}^{\log y} \frac{\cos(\delta v)}{v} dv + O\left((1 + \delta)e^{-c\sqrt{\log x}}\right), \end{aligned} \quad (3)$$

where we use the change of variable  $v = \log u$ . The error term in the first equality is evaluated as follows, taking advantage of the periodicity:

$$\left| \int_x^y \frac{\delta \sin(\delta \log u)}{u} e^{-c\sqrt{\log u}} du \right| \leq \int_{\delta \log x}^{\delta \log y} e^{-c\sqrt{w/\delta}} dw \ll \delta e^{-c\sqrt{\log x}}.$$

A similar calculation for the integral with  $\cos(\delta \log u)$  gives  $\ll e^{-c\sqrt{\log x}}$ . It remains to estimate the last integral in (3). In the case  $\delta \log y < 1$ , we can apply the approximation  $\cos(\delta v) = 1 + O(\delta^2 v^2)$  on the whole range  $\log x \leq v \leq \log y$ . This yields

$$\int_{\log x}^{\log y} \frac{\cos(\delta v)}{v} dv = \int_{\log x}^{\log y} \frac{1}{v} dv + O\left(\delta^2 \int_{\log x}^{\log y} v dv\right) = \log \log y - \log \log x + O((\delta \log y)^2),$$

as expected. For the case  $\delta \log x \geq 1$ , we integrate by parts

$$\int_{\log x}^{\log y} \frac{\cos(\delta v)}{v} dv = \frac{\sin(\delta v)}{\delta v} \Big|_{\log x}^{\log y} + \int_{\log x}^{\log y} \frac{\sin(\delta v)}{\delta v^2} dv \ll (\delta \log x)^{-1}.$$

□

## 2.2 Moments of the Steinhaus model

Consider  $(e^{i\theta(p)}, p \text{ primes})$  be IID Steinhaus random variables, i.e.,  $\theta(p)$  is uniformly distributed on  $[0, 2\pi]$ .

1. Let  $c(p)$  and  $c^*(p)$  be two sets of complex numbers indexed by the primes. Prove that for a fixed  $p$

$$\mathbf{E} \left[ \exp \left( \frac{1}{2} (c(p)e^{-i\theta(p)} + c^*(p)e^{i\theta(p)}) \right) \right] = \sum_{j \geq 0} \frac{((c(p)c^*(p))^j)}{2^{2j}(j!)^2}.$$

*Proof.* This is simply by expanding the exponential

$$\mathbf{E} \left[ \exp \left( \frac{1}{2} (c(p)e^{-i\theta(p)} + c^*(p)e^{i\theta(p)}) \right) \right] = \sum_{k \geq 0} \frac{1}{2^k k!} \sum_{\ell=0}^k \binom{k}{\ell} (c(p))^\ell (c^*(p))^{k-\ell} \mathbf{E} [e^{-i\ell\theta(p)} e^{i(k-\ell)\theta(p)}].$$

By the orthogonality relation, only the terms with  $\ell = k/2$ ,  $k$  even, survive. This leaves

$$\sum_{j \geq 0} \frac{1}{2^{2j}(j!)^2} (c(p)c^*(p))^j$$

as claimed. □

2. **MGF of Steinhaus.** Let  $z \in \mathbb{C}$  and  $1 \leq X < Y$ . Deduce from the above that that

$$\mathbf{E} \left[ \exp \left( z \sum_{X < p \leq Y} \frac{\operatorname{Re} e^{i\theta(p)}}{p^{1/2}} \right) \right] = \prod_{X < p \leq Y} I_0 \left( \frac{z}{p^{1/2}} \right), \quad (4)$$

where  $I_0$  is the zero-th order of the Bessel function of the first kind

$$I_0(w) = \sum_{k=0}^{\infty} \frac{w^{2k}}{2^{2k}(k!)^2}. \quad (5)$$

*Proof.* This is direct from Problem 1. □

3. **Gaussian domination.** Conclude that for  $\ell = 2k$  even,

$$\mathbf{E} \left[ \left( \sum_{X < p \leq Y} \frac{\operatorname{Re} e^{i\theta(p)}}{p^{1/2}} \right)^\ell \right] = \frac{(2k)!}{4^k} \sum_{\substack{p|n \Rightarrow p \in (X, Y] \\ \Omega(n) = k}} \frac{g(n)^2}{n}, \quad (6)$$

and 0 if  $\ell = 2k - 1$  is odd, and where  $g(n)$  is the multiplicative function defined by  $g(p^k) = 1/k!$ . In particular, the moments are dominated by Gaussian ones:

$$\mathbf{E} \left[ \left( \sum_{X < p \leq Y} \frac{\operatorname{Re} e^{i\theta(p)}}{p^{1/2}} \right)^{2k} \right] \leq \frac{(2k)!}{2^k k!} \left( \frac{1}{2} \sum_{X < p \leq Y} \frac{1}{p} \right)^{2k}. \quad (7)$$

*Proof.* We have from Problem 2 that

$$\begin{aligned} \prod_{X < p \leq Y} I_0 \left( \frac{z}{p^{1/2}} \right) &= \prod_{X < p \leq Y} \left( 1 + \frac{z^2}{4(1!)^2 p} + \cdots + \frac{z^{2k}}{4^k (k!)^2 p^k} + \cdots \right) \\ &= \sum_{p|n \Rightarrow p \in (X, Y]} g(n)^2 (z^2/4)^{\Omega(n)} n^{-1}, \end{aligned} \quad (8)$$

where  $\Omega(n)$  stands for the number of prime factors of  $n$  counted with multiplicities, where we noticed the Euler product form. The formula for the moments follows by identifying the coefficients on each side. The Gaussian domination follows

$$k! \sum_{\Omega(n) = k} g(n) n^{-1} = \left( \sum_p \frac{1}{p} \right)^k, \quad (9)$$

by writing the  $k$ -power as a product of  $k$  sums, and noticing that there are  $k!g(n)$  multiplicities for a given integer  $n$ . Note also that  $g(n) \leq 1$ .  $\square$

4. **Gaussian tail.** Use the above to prove a Gaussian tail estimate: for  $V \geq 1$ ,

$$\mathbf{P} \left( \sum_{X < p \leq Y} \frac{\operatorname{Re} e^{i\theta(p)}}{p^{1/2}} > V \right) \leq \exp \left( - \frac{V^2}{2s^2} \right), \quad (10)$$

where  $s^2 = \frac{1}{2} \sum_{X < p \leq Y} 1/p$ .

*Proof.* By a Chernoff bound, the probability is smaller than  $e^{-\lambda V} \mathbf{E}[\exp(\lambda \sum_{X < p \leq Y} \frac{\operatorname{Re} e^{i\theta(p)}}{p^{1/2}})]$ . But  $\mathbf{E}[\exp(\lambda \sum_{X < p \leq Y} \frac{\operatorname{Re} e^{i\theta(p)}}{p^{1/2}})] \leq e^{\lambda^2 s^2 / 2}$  by the Gaussian domination proved in the last problem. It remains to optimize by taking  $\lambda = V/s^2$ .  $\square$

### 3 Ballot Theorems

#### 3.1 Reflection Principle

1. **Why ballot?** Let  $(W_k, k \leq n)$  be a simple symmetric random walk (SSRW), i.e.,

$$W_k = W_0 + Y_1 + \cdots + Y_k$$

where  $(Y_k, k \geq 1)$  are IID with  $\mathbf{P}(Y_1 = +1) = \mathbf{P}(Y_1 = -1) = 1/2$ . Note that if  $a, b \in \mathbb{N}$  and  $a + n$  has the same parity as  $n$ ,

$$\mathbf{P}(W_n = b | W_0 = a) = \frac{1}{2^n} \binom{n}{\frac{n+b-a}{2}}.$$

- (a) Prove that for  $b \in \mathbb{N}$  (with the same parity as  $n$ )

$$\mathbf{P}\left(W_k > 0, \forall 1 \leq k \leq n, W_n = b | W_0 = 0\right) = \frac{b}{n} \mathbf{P}(W_n = b | W_0 = 0),$$

so that

$$\mathbf{P}\left(W_k > 0, \forall 1 \leq k \leq n | W_n = b, W_0 = 0\right) = \frac{b}{n}.$$

*Origin of the name "ballot":* There is an election with  $n$  votes between two candidates. Votes are counted one by one.  $W_k$  represents the difference of votes between the candidates. Given that the winner won by  $b$  votes, we are looking at the event that the eventual winner was ahead in the count at each step.

- (b) Consider now  $a, b > 0$ . Suppose for simplicity that  $a, b, n$  are all even. Prove that

$$\mathbf{P}(W_k > 0, \forall k \leq n | W_0 = a, W_n = b) \sim \frac{2ab}{n}.$$

*Proof.* (a) Let  $a, b > 0$ . The reflection principle for SSRW states the walks that start at  $a$ , end at  $b$  and go through 0 in between are in bijection with the walks starting at  $-a$  and end at  $b$ . (This is easy to see by flipping the increments from time 0 to the first time the walk reaches 0.)

Note that walks in the event considered must be such that  $W_1 = 1$ . It suffices to count the walks starting at  $W_1 = 1$  and ending at  $W_n = b$  and subtract the number of such walks that cross 0. By the reflection principle, the latter is the same as the number of walks starting at  $W_1 = -1$  and ending at  $W_n = b$ . Therefore

$$\mathbf{P}\left(W_k > 0, \forall 1 \leq k \leq n, W_n = b | W_0 = 0\right) = \frac{1}{2^n} \left( \binom{n-1}{\frac{n+b-2}{2}} - \binom{n-1}{\frac{n+b}{2}} \right) = \frac{b}{n} \frac{\binom{n}{\frac{n+b}{2}}}{2^n}.$$

- (b) The reflection principle gives

$$\mathbf{P}\left(\min_{0 \leq k \leq n} W_k \leq 0, W_n = b | W_0 = a\right) = \mathbf{P}(W_n = b | W_0 = -a).$$

Therefore, by Bayes's and the symmetry of the walk

$$\mathbf{P}\left(\min_{0 \leq k \leq n} W_k \leq 0 \mid W_0 = a, W_n = b\right) = \frac{\mathbf{P}(W_n = b \mid W_0 = -a)}{\mathbf{P}(W_n = b \mid W_0 = a)} = \frac{\mathbf{P}(W_n = -a \mid W_0 = b)}{\mathbf{P}(W_n = a \mid W_0 = b)}.$$

The ratio of these probabilities is

$$\frac{\binom{\frac{n+b+a}{2}}{\frac{n}{2}}}{\binom{\frac{n+b-a}{2}}{\frac{n}{2}}} = \frac{\binom{\frac{n-(b-a)}{2}}{\frac{n}{2}} \cdots \binom{\frac{n-(b+a)}{2} + 1}{\frac{n}{2}}}{\binom{\frac{n+(b+a)}{2}}{\frac{n}{2}} \cdots \binom{\frac{n+(b-a)}{2} + 1}{\frac{n}{2}}} = \frac{\left(1 - \frac{(b-a)}{n}\right)}{\left(1 + \frac{(b+a)}{n}\right)} \times \cdots \times \frac{\left(1 - \frac{(b+a)-2}{n}\right)}{\left(1 + \frac{(b-a)+2}{n}\right)}$$

There are  $a$  terms in this product. Using the Taylor approximation  $(1+x)^{-1} = 1-x+O(x^2)$  we get that each of these are

$$1 - \frac{2b}{n} + O(1/n^2).$$

Therefore

$$\mathbf{P}\left(\min_{0 \leq k \leq n} W_k > 0 \mid W_0 = a, W_n = b\right) = 1 - \left(1 - \frac{2b}{n} + O(1/n^2)\right)^a = \frac{2ab}{n} + O(1/n^2),$$

by the expansion  $(1+x)^a = 1+ax+O(x^2)$ . The claim follows.  $\square$

2. **Ballot theorem for Brownian motion.** Let  $(W_t, t \geq 0)$  be a standard Brownian motion. Prove that for  $a, b > 0$

$$\mathbf{P}\left(W_s \geq 0 \forall s \leq t \mid W_0 = a, W_t = b\right) = 1 - e^{-2ab/t}.$$

In particular, this shows that when  $ab = o(t)$ ,

$$\mathbf{P}\left(W_s \geq 0, \forall s \leq t \mid W_0 = a, W_t = b\right) \sim \frac{2ab}{t}. \quad (11)$$

*Proof.* Let  $I = (b - \delta, b + \delta)$  with  $\delta > 0$  be a small interval around  $b$ . We have by the *reflection principle* for Brownian motion.

$$\mathbf{P}\left(W_s \geq 0, \forall s \leq t, W_t \in I \mid W_0 = a\right) = \mathbf{P}(W_t \in I \mid W_0 = a) - \mathbf{P}(W_t \in -I \mid W_0 = a).$$

The result follows by noticing that

$$\lim_{\delta \rightarrow 0} \frac{\mathbf{P}(W_t \in -I \mid W_0 = a)}{\mathbf{P}(W_t \in I \mid W_0 = a)} = e^{-2ab/t},$$

by writing the PDF of  $W_t$  at  $b$  and  $-b$ . The above formula directly implies that for large  $t$  and  $ab = o(t)$

$$\mathbf{P}\left(W_s \geq 0, \forall s \leq t \mid W_0 = a, W_t = b\right) \sim \frac{2ab}{t}. \quad \square$$

### 3.2 Ballot theorems for Gaussian random walks.

In what follows,  $(S_k, k \leq 1)$  stands for a Gaussian walk, i.e.,

$$S_k = S_0 + Y_1 + Y_2 + \cdots + Y_k,$$

where  $(Y_k, k \geq 1)$  are independent Gaussian random variables with means 0 and variance  $s_k^2$ . The assumption that  $s_k$  is uniformly bounded away from 0 is sufficient here.

We will assume the following *Gambler's ruin* estimate: for  $a \geq 1$  and  $k \geq 10$ , we have

$$\mathbf{P}(S_j \geq 0, j \leq k | S_0 = a) \ll \frac{a}{\sqrt{k}}. \quad (12)$$

The estimate is valid not only for Gaussian increments, but for any distribution with sufficient tail decay.

1. Let  $a, b \geq 1$  and  $k \geq 10$ . Prove

$$\mathbf{P}(S_j \geq 0 \forall j \leq k | S_0 = a, S_k = b) \ll \frac{ab}{k}.$$

*Hint: Divide the interval  $[0, k]$  into three parts and use the Gambler's ruin estimate at both ends.*

*Proof.* We write for short

$$P_{(j,z)}^{(j',z')} = \mathbf{P}(S_i \geq 0 \forall j \leq i \leq j' | S_j = z, S_{j'} = z').$$

We define the times  $k_1 = \lfloor k/3 \rfloor$ ,  $k_2 = \lfloor 2k/3 \rfloor$ . The result will follow by applying a gambler's ruin estimate at the beginning and at the end of the walk and by trivially bound the middle third. More precisely, the probability can be re-expressed by integrating over the values at  $k_1$  and  $k_2$ . Writing  $f_{m,n}(x) = \frac{e^{-x^2/(2s_{m,n}^2)}}{\sqrt{2\pi s_{m,n}^2}}$  for the density, we get that the probability is

$$= \frac{1}{f_{0,k}(b-a)} \int_0^\infty \int_0^\infty P_{(0,a)}^{(k_1,x)} P_{(k_1,x)}^{(k_2,y)} P_{(k_2,y)}^{(k,b)} f_{0,k_1}(x-a) f_{k_1,k_2}(y-x) f_{k_2,k}(b-y) dx dy.$$

If  $a \leq \sqrt{k}$  and  $b \leq \sqrt{k}$ , then  $|b-a| \leq \sqrt{k}$  and we have the bound  $f_{0,k}(b-a) \gg 1/\sqrt{k}$ . Moreover, we have trivially  $P_{(k_1,x)}^{(k_2,y)} \leq 1$  and  $f_{k_1,k_2}(y-x) \ll 1/\sqrt{k}$ . We are left with

$$\ll \int_0^\infty P_{(0,a)}^{(k_1,x)} f_{0,k_1}(x-a) dx \int_0^\infty P_{(k_2,y)}^{(k,b)} f_{k_2,k}(b-y) dy \ll \frac{ab}{k},$$

by (12) applied to both integrals.

If both  $a, b$  are greater than  $\sqrt{k}$ , the estimate is trivial so we can assume  $a < \sqrt{k} < b$ . We use a standard Brownian motion  $(W_s, s \geq 0)$ . Note that  $(S_j, j \in \mathbb{N})$  has the same distribution as  $(W_{s_j^2}, j \in \mathbb{N})$ . In particular, the probability in question equals

$$\begin{aligned} & \mathbf{P}\left(W_{s_j^2} \geq 0 \forall j \leq k \mid W_0 = a, W_{s_k^2} = b\right) \\ &= \frac{\mathbf{P}\left(W_{s_j^2} \geq 0 \forall j \leq k \mid W_0 = a, W_{s_k^2} = b\right)}{\mathbf{P}\left(W_s \geq 0 \forall s \leq s_k^2 \mid W_0 = a, W_{s_k^2} = b\right)} \cdot \mathbf{P}\left(W_s \geq 0 \forall s \leq s_k^2 \mid W_0 = a, W_{s_k^2} = b\right) \end{aligned}$$

The last probability is  $\sim 2ab/k$  by (11). Now, the ratio is the inverse of the conditional probability  $\mathbf{P}\left(W_s \geq 0 \forall s \leq s_k^2 \mid W_0 = a, W_{s_k^2} = b, W_{s_j^2} \geq 0 \forall j \leq k\right)$ . This is an increasing function of  $b$ . (This is intuitively clear. A proof is given in Bramson 83 Lemma 2.5). Therefore, an upper bound of the above is obtained by replacing  $b$  by  $\sqrt{k}$  in the ratio. The numerator is then seen to be  $\ll a/\sqrt{k}$  by the last case and the denominator is  $\gg a/\sqrt{k}$  by (11). This concludes the proof.  $\square$

2. We need need a ballot estimate to hold for a non-constant barrier whose growth is less than a square root envelope, the square root the typical fluctuation for a random walk. More precisely, take a barrier  $(u_j, j \leq k)$  such that

$$|u_j| \leq (j \wedge (k - j))^\alpha, \quad 2 \leq j \leq k - 2, \quad (13)$$

and  $u_j = 0$  for  $j = 0, 1, k - 1, k$ .

Let  $a, b \geq 1, k \geq 10$  and  $0 < \alpha < 1/2$ . Consider  $(u_j, j \leq k)$  as in (13). Prove that

$$\mathbf{P}(S_j \geq -u_j \forall j \leq k \mid S_0 = a, S_k = b) \ll \frac{ab}{k}.$$

*Proof.* Without loss of generality, we can assume that  $ab \leq k$ . We bound the worst case when  $u_j = (j \wedge (k - j))^\alpha$ . We write  $\Delta u_j = u_j - u_{j-1}$  for the increments of the barrier and  $\Delta S_j = S_j - S_{j-1}$  for the increments of the walk. We can write the probability in terms of the distribution of the increments as

$$\frac{1}{f_k(b)} \int \mathbf{1}\left(a + \sum_{i=1}^j y_i \geq -u_j \forall j; \sum_{i=1}^k y_i = b - a\right) \prod_j \frac{e^{-y_j^2/(2\Delta s_j^2)}}{\sqrt{2\pi\Delta s_j^2}} dy_j,$$

where  $f_k(b)$  is the PDF of  $S_k$  at  $b$ . We make the change of variable  $y'_j = y_j + \Delta u_j$  to eliminate the drift, and bound the factor  $e^{-\sum_j \Delta u_j^2/(2\Delta s_j^2)}$  by 1, since the sum is small by definition of the  $u_j$ 's. We get the bound

$$\ll \mathbf{E} \left[ e^{\sum_{j=1}^k 2\Delta u_j \Delta S_j / (2\Delta s_j^2)} \mathbf{1}(S_j \geq 0 \forall j \leq k) \mid S_0 = a, S_k = b \right].$$

It is convenient to define the walk recentered by the linear interpolation of the boundary values, that is,  $\overline{S}_j = S_j - a(1 - \frac{j}{k}) - b\frac{j}{k}$  and the corresponding increment  $\Delta\overline{S}_j$ . The sum in the exponential factor is then re-expressed as

$$\sum_{j=1}^k \frac{\Delta u_j \Delta\overline{S}_j}{\Delta s_j^2} + \sum_{j=1}^k \frac{\Delta u_j}{\Delta s_j^2} \frac{b-a}{k}.$$

The second sum is  $\ll \sum_{j=1}^k \frac{\Delta u_j}{\Delta s_j^2} \ll \sum_{j=1}^k e^{-j} |\Delta u_j| \ll 1$  since  $ab \leq k$  and  $\sum_{j=1}^k \Delta u_j = 0$ .

The first sum can be rewritten by parts as  $\sum_{j=1}^{k-1} \alpha_j \overline{S}_j$  with  $-\alpha_j = \frac{\Delta u_{j+1}}{\Delta s_{j+1}^2} - \frac{\Delta u_j}{\Delta s_j^2}$ . (Here we used the fact that  $\Delta u_1 = 0$  and  $\Delta u_k = 0$  by definition of the  $u_j$ 's to eliminate the boundary terms.) We are left with evaluating

$$\mathbf{E} \left[ e^{\sum_{j=1}^{k/2} \alpha_j \Delta\overline{S}_j} \mathbf{1}(S_j \geq 0 \forall j \leq k) \middle| S_0 = a, S_k = b \right],$$

where we used the inequality  $AB \leq A^2/2 + B^2/2$  to reduce the summation to  $j \leq k/2$ . We partition over the possible values of the summand. For values smaller than 1 (say), we can use the bound of the linear barrier. For values larger than 1, we first note that  $0 < a_j < cj^{\alpha-2}$  for some  $c > 0$  by Taylor's theorem applied to  $u_j$ . This implies that if  $\sum_{j=1}^{k/2} a_j \Delta\overline{S}_j \in [v, v+1]$  for some  $v \geq 1$ , then there must exist  $\ell \leq k/2$  with  $\overline{S}_\ell > v\ell^{1/2+\varepsilon}$ , where  $\varepsilon$  is chosen smaller than  $1/2 - \alpha$ . By a union bound on  $\ell$  and  $v$ , we get the following upper bound by conditioning on  $\overline{S}_\ell$

$$\sum_{\ell=1}^{k/2} \sum_{v \geq 1} e^v \int_{v\ell^{1/2+\varepsilon}}^{\infty} \mathbf{P}(S_j \geq 0 \forall j \leq k \mid S_0 = a, S_k = b, \overline{S}_\ell = w) g_\ell(w) dw,$$

where  $g_\ell(w)$  is the PDF of  $\overline{S}_\ell$  under  $\mathbf{P}(\cdot \mid S_0 = a, S_k = b)$ . By definition,  $\overline{S}_\ell$  has mean 0 under this measure and variance  $s_\ell^2 - \frac{\ell}{k} s_k^2 \leq \ell$ . We get an upper bound by dropping  $\{S_j \geq 0\}$  for  $j < \ell$ . The probability can then be evaluated using a linear barrier with starting point  $w + a(1 - \frac{\ell}{k}) + b\frac{\ell}{k}$  at time  $\ell$ . Put together, this gives the bound

$$\ll \frac{1}{k} \sum_{\ell=1}^{k/2} \sum_{v \geq 1} e^v \int_{v\ell^{1/2+\varepsilon}}^{\infty} b(w + a(1 - \frac{\ell}{k}) + b\frac{\ell}{k}) e^{-w^2/(2\ell)} dw.$$

The integral is  $\ll b(w + a(1 - \frac{\ell}{k}) + b\frac{\ell}{k}) \sqrt{\ell} e^{-v^2 \ell^{2\varepsilon}/2}$ . The sum over  $v$  is then dominated by the term at  $\ell = 1$ ,  $e^{-\ell^{2\varepsilon}/2}$ . The last sum over  $\ell$  finally yields the claimed bound  $ab/k$ .  $\square$

## 4 Gaussian Approximation and Harmonic Analysis

### 4.1 Berry-Esseen Estimates

1. **Approximating indicator functions by smooth functions.** Let  $a < b$  and  $0 < \delta < \min\{1, \frac{b-a}{2}\}$ . Construct a function  $g$  depending on  $a, b, \delta$  in the Schwartz space of  $C^\infty$  functions with rapid decay such that:

- (i)  $0 \leq g(x) \leq 1$ ,  $x \in \mathbb{R}$ ,
- (ii) its Fourier transform satisfies  $|\hat{g}(\xi)| \leq e^{-\pi^2 \delta^4 \xi^2}$ ,  $\xi \in \mathbb{R}$ ,
- (iii) for any  $x \in \mathbb{R}$ ,  $(1 - e^{-1/\delta^2}) \mathbf{1}_{[a,b]}(x) \leq g(x) \leq \mathbf{1}_{[a-2\delta, b+2\delta]}(x) + e^{-1/\delta^2}$ .

*Hint:* Convolve with the Gaussian kernel  $\varphi_c(y) = \frac{e^{-y^2/c^2}}{\sqrt{\pi c^2}}$  with  $c < \delta$ .

*Proof.* We consider the Gaussian kernel  $\varphi_c(y) = \frac{e^{-y^2/c^2}}{\sqrt{\pi c^2}}$  with  $c < \delta$ . We choose  $c = \delta^2$ . The approximation of the indicator function  $\mathbf{1}_{[a,b]}$  is the convolution of  $\varphi_c$  with an indicator of a slightly larger interval

$$g(x) = \varphi_c * \mathbf{1}_{[a-\delta, b+\delta]}(x) = \int_{\mathbb{R}} \mathbf{1}_{[a-\delta, b+\delta]}(y) \frac{e^{-(x-y)^2/c^2}}{\sqrt{\pi c^2}} dy.$$

This function obviously satisfies (i). Moreover, the Fourier transform of  $\mathbf{1}_{[a-\delta, b+\delta]}$  and of  $\varphi_c$  are by a direct computation

$$\hat{\mathbf{1}}_{[a-\delta, b+\delta]}(\xi) = \int_{a-\delta}^{b+\delta} e^{-i2\pi\xi x} dx = e^{-i\pi\xi(a+b)} \frac{\sin(\pi\xi(b-a+2\delta))}{\pi\xi},$$

and  $\hat{\varphi}_c(\xi) = e^{-\pi^2 \xi^2 c^2}$ . In particular, the bound  $|\hat{g}(\xi)| = |\hat{\mathbf{1}}_{[a-\delta, b+\delta]}(\xi) \hat{\varphi}_c(\xi)| \leq e^{-\pi^2 \xi^2 \delta^4}$  is satisfied.

It remains to prove (iii). We prove the first inequality for  $g$ . The second follows by a similar argument. The inequality is obvious for  $x \notin [a, b]$ . For  $x \in [a, b]$ , the following bound holds

$$g(x) = \int_{a-\delta}^{b+\delta} \frac{e^{-(y-x)^2/c^2}}{\sqrt{\pi c^2}} dy \geq 1 - 2 \int_{\delta}^{\infty} \frac{e^{-y^2/c^2}}{\sqrt{\pi c^2}} dy \geq 1 - e^{-\delta^2/c^2}.$$

This is the claimed bound by the choice of  $c$ . □

2. **Berry-Esseen for Steinhaus.** Let  $x < y$  and  $a < b$ . Consider the Steinhaus model

$$S_{x,y}(h) = \sum_{e^x < \log p \leq e^y} \frac{\operatorname{Re} e^{i\theta(p)} p^{-ih}}{p^{1/2}} \quad (14)$$

and the Gaussian model

$$\mathcal{N}_{x,y}(h) = \sum_{e^x < \log p \leq e^y} \frac{\operatorname{Re} Z(p) p^{-ih}}{p^{1/2}} \quad (15)$$

where  $(Z(p), p \text{ primes})$  are IID standard complex Gaussian random variables.

Use the previous question to show that for  $h \in \mathbb{R}$

$$\mathbf{P}(S_{x,y}(h) \in [a, b]) = \mathbf{P}(\mathcal{N}_{x,y}(h) \in [a, b]) + O\left(e^{-e^x/100}\right),$$

uniformly in  $a, b, x, y$ .

*Proof.* We set  $X = \exp(e^x)$  and drop the dependence on  $h$  throughout the proof for simplicity. Consider the approximate indicator function  $g$  as in the previous question. We have that

$$\mathbf{P}(S_{x,y} \in [a, b]) \leq \mathbf{E}[g(S_{x,y})] + O(e^{-1/\delta^2}).$$

We choose  $\delta = X^{-1/32}$ . We now prove that

$$\mathbf{E}[g(S_{x,y})] = \mathbf{E}[g(\mathcal{N}_{x,y})] + O(X^{-1/2}). \quad (16)$$

In Fourier space, the first expectation is expressed as

$$\mathbf{E}[g(S_{x,y})] = \int_{\mathbb{R}} \hat{g}(\xi) \mathbf{E}[e^{2\pi i \xi S_{x,y}}] d\xi.$$

It remains to control the integral on the different ranges of the Fourier modes. For  $|\xi| \leq X^{1/8}$ , Equation (4) with  $z = 2\pi\xi$  and  $z' = 0$  implies

$$\mathbf{E}[e^{2\pi i \xi S_{x,y}}] = \mathbf{E}[e^{2\pi i \xi \mathcal{N}_{x,y}}] \left(1 + O(X^{-1/2})\right) = \mathbf{E}[e^{2\pi i \xi \mathcal{N}_{x,y}}] + O(X^{-1/2}).$$

For  $|\xi| > X^{1/8}$ , the bound (ii) on the Fourier transform of  $g$  gives

$$\mathbf{E}[g(S_{x,y})] \leq \int_{|\xi| > X^{1/8}} |\hat{g}(\xi)| d\xi \leq \int_{|\xi| > X^{1/8}} e^{-\pi^2 \delta^4 \xi^2} d\xi \ll e^{-\pi^2 X^{1/4} \delta^4},$$

which is easily  $O(X^{-1/2})$ . This proves the claim (16). Applying the approximation of  $g$  this time to  $g(\mathcal{N}_{x,y})$  yields for  $\delta = X^{-1/32}$

$$\mathbf{P}(S_{x,y} \in [a, b]) \leq \mathbf{P}(\mathcal{N}_{x,y} \in [a - 2\delta, b + 2\delta]) + O(X^{-1/2}) = \mathbf{P}(\mathcal{N}_{x,y} \in [a, b]) + O(X^{-1/32}),$$

by simply bound the probability of the overflows  $(b, b + 2\delta]$  and  $[a - 2\delta, a)$  by  $2\delta$ . The reverse inequality is proved the same way by applying the same approximation with  $[a, b]$  replaced by  $[a + 2\delta, b - 2\delta]$ .  $\square$

3. **Gaussian Decoupling.** For  $|\rho| \leq 1$  and  $\mathfrak{s}^2 > 0$ . consider the two-dimensional Gaussian vectors  $\mathbf{X}$  and  $\mathbf{X}'$  with mean 0 and respective covariances

$$\mathcal{C} = \mathfrak{s}^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \mathcal{C}' = \mathfrak{s}^2 \begin{pmatrix} 1 + |\rho| & 0 \\ 0 & 1 + |\rho| \end{pmatrix}.$$

Then for any Borel measurable set  $A \in \mathbb{R}^2$ , we have

$$\mathbf{P}(\mathbf{X} \in A) \leq \left( \frac{1 + |\rho|}{1 - |\rho|} \right)^{1/2} \cdot \mathbf{P}(\mathbf{X}' \in A)$$

*Proof.* By rescaling  $\mathbf{X}$  and  $\mathbf{X}'$  by  $1/\mathfrak{s}$ , we can simply assume that  $\mathfrak{s}^2 = 1$ . Let's first prove the case  $\rho \geq 0$ . The result simply follows by expanding the PDF of  $\mathbf{X}$ : for  $\mathbf{x}^T = (x_1, x_2) \in \mathbb{R}^2$ , we have

$$\begin{aligned} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathcal{C}^{-1} \mathbf{x}\right) &= \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right) \\ &\leq \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right) \\ &\leq \left(\frac{1+\rho}{1-\rho}\right)^{1/2} \frac{1}{(1+\rho)^{1/2}} \exp\left(-\frac{x_1^2 + x_2^2}{2(1+\rho)}\right), \end{aligned}$$

since  $x_1^2 - 2\rho x_1 x_2 + x_2^2 \geq (1-\rho)(x_1^2 + x_2^2) \geq 0$  for  $\rho \geq 0$ . This proves the claim for  $\rho \geq 0$ . If  $\mathbf{X}^T = (X_1, X_2)$  is such that  $\rho < 0$ , it suffices to consider  $(X_1, -X_2)$  and apply the result above.  $\square$

## 4.2 Discretization

1. **via Chaining.** Consider the Steinhaus model. Let  $V > 0$ . Then we have for some absolute constant  $c$

$$\mathbf{P}\left(\max_{|h| \leq e^{-k}} \{S_k(h) - S_k(0)\} > V\right) \ll e^{-cV^2}.$$

*Hint:* Use a chaining argument. Consider a discretization of the intervals with dyadic mesh size

$$H_\ell = [-e^{-k}, e^{-k}] \cap \frac{e^{-k}\mathbb{Z}}{2^\ell}, \ell \geq 0.$$

Then write

$$S_k(h) - S_k(0) = \sum_{\ell \geq 0} S_k(h_{\ell+1}) - S_k(h_\ell),$$

where the chain  $h_\ell$ ,  $\ell \geq 0$ , is such that  $h_0 = 0$ ,  $|h_{\ell+1} - h_\ell| \leq e^{-k}/2^{\ell+1}$ , and  $h_\ell \rightarrow h$ .

*Proof.* Fix  $k \geq 1$ . Note that  $\#H_\ell \leq 2^{\ell+1} + 1$ . Now, let  $(d_\ell, \ell \geq 0)$  be a sequence of positive numbers with  $\sum_{\ell \geq 0} d_\ell = 1$ . There is some freedom in the choice of  $d_\ell$  but it should not decay too fast as we will see later. We pick  $d_\ell = \frac{6}{\pi^2} \frac{1}{(1+\ell)^2}$ . If  $S_k(h) - S_k(0) > V$  for some  $h$ , then we must have that  $S_k(h_{\ell+1}) - S_k(h_\ell) > d_\ell V$  for some  $\ell$ . The probability can then be bounded by a (massive) union bound on  $\ell$  and  $H_\ell$

$$\ll \sum_{\ell \geq 0} \sum_{v \in H_\ell} \mathbf{P}(S_k(v^+) - S_k(v) \geq d_\ell V), \quad (17)$$

where for a given  $v \in H_\ell$ ,  $v^+$  denote the closest point in  $\mathcal{H}_{\ell+1}$  not equal to  $v$ . (There are two such points, but that multiplicity is absorbed in  $\ll$ .) The inequality  $|v^+ - v| \leq$

$e^{-k}/2^{\ell+1}$  holds by definition. It remains to apply the Gaussian tail estimate (10) with  $a(p) = p^{-iv^+} - p^{-iv}$  since

$$S_k(v^+) - S_k(v) = \sum_{p \leq \exp(e^k)} \frac{\operatorname{Re}((p^{-iv^+} - p^{-iv})e^{i\theta(p)})}{p^{1/2}}.$$

The variance of this sum is for a given  $v$

$$\mathfrak{s}_k(v)^2 = \sum_{p \leq \exp(e^k)} \frac{1 - \cos(|v^+ - v| \log p)}{p} \leq \sum_{p \leq \exp(e^k)} \frac{(|v^+ - v| \log p)^2}{p} \leq 2^{-2(\ell+1)}. \quad (18)$$

The first inequality follows from the fact that  $1 - \cos x \leq x^2$  for  $x \leq 1/2$  and  $|v^+ - v| \log p \leq 2^{-(\ell+1)}$ . We use Mertens's estimate in the second. This then yields

$$\ll \sum_{\ell \geq 0} 2^\ell \exp\left(\frac{-d_\ell^2 V^2}{2\mathfrak{s}_\ell^2}\right) \leq \sum_{\ell \geq 0} 2^\ell \exp\left(-2^\ell d_\ell^2 V^2\right).$$

We see that it suffices to take the sequence  $d_\ell$  so that the above sum is convergent. The above sum is then  $\ll \exp(-cV^2)$  for some absolute constant as claimed.  $\square$

2. **via Poisson summation.** Let  $\varepsilon > 0$  be given. Let  $V$  be a smooth function (say in the Schwartz class) with  $V(x) = 1$  for  $0 \leq x \leq 1$  and supported on  $[-1, 2]$  (say). We write  $\widehat{V}$  for its Fourier transform.

(a) Use Poisson summation to show that for any positive integer  $n \leq N$  and any  $h \in \mathbb{R}$

$$n^{-ih_0} = \frac{1}{3} \sum_{h \in \frac{2\pi\mathbb{Z}}{3 \log N}} n^{-ih} \widehat{V}\left(\frac{(h - h_0) \log N}{2\pi}\right).$$

*Hint: Consider  $\overline{G}(x) = V(2\pi x / \log N)$  and  $\sum_{k \in \mathbb{Z}} f(k)$  with*

$$f(x) = n^{-\frac{2\pi ix}{3 \log N}} \widehat{G}\left(\frac{2\pi x}{3 \log N} - h_0\right)$$

(b) Conclude that for a Dirichlet polynomial  $D(h)$  of the form  $\sum_{n \leq N} a_n n^{-i(t+h)}$  its Steinhaus equivalent  $\sum_{n \leq N} a_n e^{i\theta(n)} n^{-ih}$ , we have

$$\max_{|h| \leq 1} |D(h)|^2 \ll \sum_{|j| \leq \log N} \left| D\left(\frac{2\pi ij}{3 \log N}\right) \right|^2 + \sum_{|j| > \log N} \frac{1}{1 + |j|^A} \left| D\left(\frac{2\pi ij}{3 \log N}\right) \right|^2, \quad (19)$$

for  $A = 100$  (say).

*Proof.* (a) Take  $\varepsilon = 1$  for simplicity. Note that  $\widehat{G}(x) = \frac{\log N}{2\pi} \widehat{V}\left(\frac{x \log N}{2\pi}\right)$ . By Poisson summation, we have

$$\sum_{k \in \mathbb{Z}} n^{-\frac{2\pi i k}{3 \log N}} \widehat{G}\left(\frac{2\pi k}{3 \log N} - h_0\right) = \sum_{\ell \in \mathbb{Z}} \int_{\mathbb{R}} n^{-\frac{2\pi i x}{3 \log N}} \widehat{G}\left(\frac{2\pi x}{3 \log N} - h_0\right) e^{-2\pi i \ell x} dx. \quad (20)$$

For fixed  $\ell$ , taking the inverse Fourier transform, the above integral is

$$\begin{aligned} & \frac{3 \log N}{2\pi} \int_{\mathbb{R}} e^{-iy(\log n + 3\ell \log N)} \widehat{G}(y - h_0) dy \\ &= \frac{3 \log N}{2\pi} e^{-ih_0(\log n + 3\ell \log N)} G\left(-\frac{\log n + 3\ell \log N}{2\pi}\right). \end{aligned}$$

From the compact support assumption on  $V$ , for  $0 \leq n \leq N$  the above right-hand side is nonzero only for  $\ell = 0$ . Equation (20) can therefore be written as

$$n^{-ih_0} = \frac{1}{2 + \varepsilon} \sum_{h \in \frac{2\pi\mathbb{Z}}{(2+\varepsilon)\log N}} n^{-ih} \widehat{V}\left(\frac{(h - h_0) \log N}{2\pi}\right).$$

□