LECTURE NOTES 1 FOR CAMBRIDGE PART III COURSE ON "PROBABILISTIC NUMBER THEORY", MICHAELMAS 2015

ADAM J HARPER

Abstract. These are rough notes covering the first block of lectures in the "Probabilistic Number Theory" course. In these first lectures we will introduce the class of additive functions, estimate their low moments and present applications to bilinear sums, and estimate all moments to prove the Erdős–Kac central limit theorem. Throughout we emphasise the nature of all these results as statements about sums of "almost independent" random variables.

(No originality is claimed for any of the contents of these notes, which borrow from the *Probabilistic Number Theory* books of Elliott as well as from numerous original research papers.)

1. Introduction to additive functions

We begin with a simple definition that lies at the heart of classical probabilistic number theory.

Definition 1.1. A function $f : \mathbb{N} \to \mathbb{C}$ is said to be *additive* if

$$f(ab) = f(a) + f(b) \quad \forall (a, b) = 1.$$

If moreover

$$f(ab) = f(a) + f(b) \quad \forall a, b,$$

then f is said to be *completely additive*.

Notice that if $n = p_1^{a_1} ... p_k^{a_k}$, where the p_i are distinct primes and $a_i \in \mathbb{N}$, then for an additive function f we have

$$f(n) = \sum_{i=1}^{k} f(p_i^{a_i}).$$

Thus the values of f on all natural numbers are determined by its values on prime powers, so we see that additive functions might be able to pick out interesting number theoretic structure. Two key examples, which we shall discuss extensively later, are the functions

$$\omega(n) := \sum_{p|n} 1 \quad \text{and} \quad \Omega(n) := \sum_{p^k|n} 1,$$

counting the number of prime factors of n without and with multiplicity.

Date: 30th October 2015.

Probabilistic number theory began with the (gradual) observation that additive functions f are also related to some interesting probabilistic structure, at least heuristically (and, as it turns out, also rigorously). If $n \leq N$ for some fixed $N \in \mathbb{N}$, then we can write

$$f(n) = \sum_{p^k || n} f(p^k) = \sum_{p \leq N} \sum_{k=1}^{\infty} f(p^k) \mathbf{1}_{p^k || n} =: \sum_{p \leq N} f_p(n),$$

say, where $\mathbf{1}$ denotes the indicator function. Now let us consider the behaviour of the values $f(n)$ as $n \leq N$ varies, which is equivalent to considering $f(n)$ as a random variable on the probability space $([N], \mathcal{P}([N]), \mathbb{P}_N)$, where $\mathbb{P}_N$ denotes the discrete uniform measure on $[N]$. We have already expressed f as a sum of the random variables f_p, but how do the f_p behave?

We can note that, if $p_1, ..., p_w$ are any given distinct primes and $k_1, ..., k_w$ are any natural numbers, then

$$\begin{aligned} \mathbb{P}_N(p_i^{k_i} || n \; \forall 1 \leq i \leq w) &= \frac{1}{N} \#\{n \leq N : p_i^{k_i} \mid n \; \forall 1 \leq i \leq w, \text{ but } p_i^{k_i+1} \nmid n \; \forall 1 \leq i \leq w\} \\ &= \frac{1}{N} \left(\left\lfloor \frac{N}{p_1^{k_1+1} ... p_w^{k_w+1}} \right\rfloor (p_1 - 1)...(p_w - 1) + O(p_1 ... p_w) \right), \end{aligned}$$

since in each complete set of residue classes mod $p_1^{k_1+1}...p_w^{k_w+1}$, precisely $(p_1-1)...(p_w-1)$ classes will be divisible by all the $p_i^{k_i}$ but none of the $p_i^{k_i+1}$. Since we always have $\lfloor x \rfloor = x + O(1)$, it follows that

$$\mathbb{P}_N(p_i^{k_i} || n \; \forall 1 \leq i \leq w) = \prod_{i=1}^{w} \frac{(p_i - 1)}{p_i^{k_i+1}} + O(\frac{p_1 ... p_w}{N}).$$

In other words, provided that $\frac{p_1 ... p_w}{N}$ is "small", the probability will roughly resemble the product $\prod_{i=1}^{w} \frac{(p_i-1)}{p_i^{k_i+1}}$, implying that *the events $p_i^{k_i} || n$ are roughly independent for different primes p_i*. Since the random variables f_p are solely determined by events of the form $p^k || n$, it follows (as was perhaps first appreciated by Kac) that *an additive function f is a sum of roughly independent random variables f_p.*

We know lots about the behaviour of sums of independent random variables, and a theme of modern probability is to understand how this extends to various kinds of approximate independence. This is exactly the situation we are in when studying additive functions.

We end this first section by proving a simple estimate for the first moment (i.e. the mean) of an additive function, mimicking the calculation above.

Lemma 1.2. *Let f be an additive function, let $N \in \mathbb{N}$, and let $\mathbb{E}_N$ denote expectation with respect to the discrete uniform measure on $[N]$. Then*

$$\mathbb{E}_N f = \sum_{p^k \leq N} \frac{f(p^k)}{p^k}\left(1 - \frac{1}{p}\right) + O\left(\frac{1}{N}\sum_{p^k \leq N} |f(p^k)|\right).$$

In the special case of the prime factor functions ω and Ω, we have

$$\mathbb{E}_N \omega = \mathbb{E}_N \Omega + O(1) = \sum_{p \leq N} \frac{1}{p} + O(1) = \log\log N + O(1).$$

Proof of Lemma 1.2. By linearity of expectation, we have

$$\mathbb{E}_N f = \sum_{p \leq N} \mathbb{E}_N f_p = \sum_{p \leq N} \sum_{k : p^k \leq N} f(p^k) \mathbb{E}_N(\mathbf{1}_{p^k || n}) = \sum_{p \leq N} \sum_{k : p^k \leq N} f(p^k) \mathbb{P}_N(p^k || n),$$

noting that if $p^k > N$ then we do not have $p^k || n$ for any $n \leq N$. Cutting off the sums as soon as $p^k > N$ means there are no convergence issues here. Then by definition of $\mathbb{P}_N$, we have

$$\mathbb{E}_N f = \sum_{p \leq N} \sum_{k : p^k \leq N} f(p^k) \frac{1}{N}\left(\lfloor \frac{N}{p^k} \rfloor - \lfloor \frac{N}{p^{k+1}} \rfloor\right) = \sum_{p^k \leq N} f(p^k)\left(\frac{1}{p^k} - \frac{1}{p^{k+1}}\right) + O\left(\frac{1}{N}\sum_{p^k \leq N} |f(p^k)|\right).$$

This is a trivial rewriting of the estimate claimed in Lemma 1.2.

When $f = \omega$ we have $f(p^k) = 1$ for all primes p and natural numbers k. Inserting these values in the general statement we just proved, we find

$$\mathbb{E}_N \omega = \sum_{p \leq N} \frac{1}{p}\left(1 - \frac{1}{p}\right) + O\left(\sum_{p^k \leq N : k \geq 2} \frac{1}{p^k}\right) + O\left(\frac{1}{N}\sum_{p^k \leq N} 1\right).$$

The final "big Oh" term here is certainly $\ll 1$ (in fact it is $\ll \frac{1}{\log N}$), and the first "big Oh" term is $\ll \sum_{p^2 \leq N} \frac{1}{p^2} \ll 1$. Removing the subtracted terms $-1/p^2$ from the main term may similarly be done at a cost of $O(1)$, giving the result claimed in the lemma. When $f = \Omega$ we have $f(p^k) = k$, but the same argument applies because this only differs from ω on the higher prime powers in the error terms, where additional factors of k make no difference (e.g. to the convergence of geometric series). □

2. The Turán–Kubilius inequality for the variance

In Lemma 1.2 we gave an estimate for the mean $\mathbb{E}_N f$ of an additive function, and in this section we shall use similar calculations to obtain a famous estimate for the variance. As we shall see, although the proof is fairly straightforward this inequality has powerful consequences.

Theorem 2.1 (Turán–Kubilius inequality)**.** *Let f be an additive function, let $N \in \mathbb{N}$, and let $\mathbb{E}_N$ denote expectation with respect to the discrete uniform measure on $[N]$. Then*

$$\mathbb{E}_N |f - \mathbb{E}_N f|^2 \ll \sum_{p^k \leq N} \frac{|f(p^k)|^2}{p^k}.$$

Turán proved the above inequality in 1934 in the special case where $f = \omega$, and later he handled some more general cases. The fully general statement we have given (roughly speaking) was formulated by Kubilius in 1964.

Proof of Theorem 2.1. Notice first that if f is a complex-valued additive function, and we write $f = f_1 + if_2$, then the real and imaginary parts f_1, f_2 will be real-valued additive functions. Moreover we have

$$\mathbb{E}_N |f - \mathbb{E}_N f|^2 = \mathbb{E}_N |(f_1 - \mathbb{E}_N f_1) + i(f_2 - \mathbb{E}_N f_2)|^2 = \mathbb{E}_N |f_1 - \mathbb{E}_N f_1|^2 + \mathbb{E}_N |f_2 - \mathbb{E}_N f_2|^2,$$

so we see it will suffice to prove the Turán–Kubilius inequality for f_1, f_2 separately. Therefore we shall assume henceforth that f is real-valued.

For real-valued f, expanding out the bracket yields as usual that

$$\mathbb{E}_N (f - \mathbb{E}_N f)^2 = \mathbb{E}_N (f^2 - 2f\mathbb{E}_N f + (\mathbb{E}_N f)^2) = \mathbb{E}_N (f^2) - (\mathbb{E}_N f)^2.$$

Similarly as in the proof of Lemma 1.2, inserting the definition of f gives that

$$\mathbb{E}_N f^2 = \mathbb{E}_N \left(\sum_{p^k \leq N} f(p^k) \mathbf{1}_{p^k || n} \right)^2 = \sum_{p^k, q^l \leq N} f(p^k) f(q^l) \mathbb{E}_N (\mathbf{1}_{p^k || n} \mathbf{1}_{q^l || n}).$$

We shall give a different treatment of those terms in the sum where $p = q$ (and so the events $p^k || n$ and $q^l || n$ will *not* be approximately independent), and those terms where $p \neq q$ (and so the events will be approximately independent). If $p = q$ but $k \neq l$ then the events $p^k || n$ and $p^l || n$ cannot occur simultaneously, and so those summands give a zero contribution. So we can write

$$\mathbb{E}_N f^2 = \sum_{p^k \leq N} f(p^k)^2 \mathbb{E}_N (\mathbf{1}_{p^k || n}) + \sum_{\substack{p^k, q^l \leq N, \\ p \neq q}} f(p^k) f(q^l) \mathbb{E}_N (\mathbf{1}_{p^k || n} \mathbf{1}_{q^l || n}).$$

Now inserting the definition of $\mathbb{E}_N$, when $p \neq q$ we have

$$\begin{aligned} \mathbb{E}_N (\mathbf{1}_{p^k || n} \mathbf{1}_{q^l || n}) = \mathbb{P}_N (p^k || n \text{ and } q^l || n) &= \frac{1}{N} \left(\left\lfloor \frac{N}{p^k q^l} \right\rfloor - \left\lfloor \frac{N}{p^{k+1} q^l} \right\rfloor - \left\lfloor \frac{N}{p^k q^{l+1}} \right\rfloor + \left\lfloor \frac{N}{p^{k+1} q^{l+1}} \right\rfloor \right) \\ &= \left(\frac{1}{p^k} - \frac{1}{p^{k+1}} \right) \left(\frac{1}{q^l} - \frac{1}{q^{l+1}} \right) + O(1/N) \\ &= \mathbb{E}_N (\mathbf{1}_{p^k || n}) \mathbb{E}_N (\mathbf{1}_{q^l || n}) + O(1/N). \end{aligned}$$

This estimate is fairly good (in the sense that the "big Oh" term is smaller than the main term) when the product $p^k q^l \leq N$, but otherwise it is poor. Indeed, when $p^k q^l > N$

we see that the events $p^k||n$ and $q^l||n$ can in fact never occur simultaneously (and so $\mathbb{E}_N(\mathbf{1}_{p^k||n}\mathbf{1}_{q^l||n}) = 0$), because they require at least that $p^kq^l|n$. Therefore we really always have

$$\mathbb{E}_N(\mathbf{1}_{p^k||n}\mathbf{1}_{q^l||n}) = \mathbb{E}_N(\mathbf{1}_{p^k||n})\mathbb{E}_N(\mathbf{1}_{q^l||n}) + O(\min\{1/N, 1/(p^kq^l)\}).$$

Putting everything together, we have shown that

$$\begin{aligned}\mathbb{E}_N f^2 &= \sum_{p^k\le N} f(p^k)^2\mathbb{E}_N(\mathbf{1}_{p^k||n}) + \sum_{\substack{p^k,q^l\le N,\\ p\neq q}} f(p^k)f(q^l)(\mathbb{E}_N(\mathbf{1}_{p^k||n})\mathbb{E}_N(\mathbf{1}_{q^l||n}) + O(\min\{\frac{1}{N}, \frac{1}{p^kq^l}\}))\\ &= \left(\sum_{p^k\le N} f(p^k)\mathbb{E}_N(\mathbf{1}_{p^k||n})\right)^2 + \sum_{p^k\le N} f(p^k)^2(\mathbb{E}_N(\mathbf{1}_{p^k||n}) - \mathbb{E}_N(\mathbf{1}_{p^k||n})^2)\\ &\quad - \sum_{\substack{p^k,p^l\le N,\\ k\neq l}} f(p^k)f(p^l)\mathbb{E}_N(\mathbf{1}_{p^k||n})\mathbb{E}_N(\mathbf{1}_{p^l||n}) + O\left(\sum_{p^k,q^l\le N} |f(p^k)||f(q^l)|\min\{\frac{1}{N}, \frac{1}{p^kq^l}\}\right).\end{aligned}$$

Using the trivial inequality $|f(p^k)||f(q^l)| \le |f(p^k)|^2 + |f(q^l)|^2$, as well as the symmetry between p^k and q^l, it is easy to see that the contribution to the "big Oh" term from those products $p^kq^l \le N$ is

$$\ll \frac{1}{N}\sum_{p^kq^l\le N} |f(p^k)|^2 \ll \frac{1}{N}\sum_{p^k\le N} |f(p^k)|^2\frac{N}{p^k} \ll \sum_{p^k\le N}\frac{|f(p^k)|^2}{p^k},$$

which is acceptable for Theorem 2.1. The contribution from those $p^kq^l > N$ will be handled similarly, but with an extra technical trick of using the inequality

$$|f(p^k)||f(q^l)| = \frac{|f(p^k)||f(q^l)|\sqrt{\log(p^k)\log(q^l)}}{\sqrt{\log(p^k)\log(q^l)}} \le \frac{|f(p^k)|^2\log(q^l)}{\log(p^k)} + \frac{|f(q^l)|^2\log(p^k)}{\log(q^l)}.$$

This implies that

$$\sum_{\substack{p^k,q^l\le N,\\ p^kq^l>N}} \frac{|f(p^k)||f(q^l)|}{p^kq^l} \ll \sum_{p^k\le N}\frac{|f(p^k)|^2}{p^k\log(p^k)}\sum_{N/p^k\le q^l\le N}\frac{\log(q^l)}{q^l} \ll \sum_{p^k\le N}\frac{|f(p^k)|^2}{p^k\log(p^k)}\log(p^k),$$

which again is acceptable. Easy arguments show that the second and third sums in our expression for $\mathbb{E}_N f^2$ are also $\ll \sum_{p^k\le N}\frac{|f(p^k)|^2}{p^k}$, and we have $\left(\sum_{p^k\le N} f(p^k)\mathbb{E}_N(\mathbf{1}_{p^k||n})\right)^2 = (\mathbb{E}_N f)^2$, so Theorem 2.1 follows. □

We give two immediate corollaries in our favourite special case of the prime divisor functions ω, Ω.

Corollary 2.2. *We have*

$$\mathbb{E}_N(\omega - \log\log N)^2 \ll \log\log N \quad \textit{and} \quad \mathbb{E}_N(\Omega - \log\log N)^2 \ll \log\log N.$$

Corollary 2.3. *Given any function $t(N) \geq 1$, we have*

$$\mathbb{P}_N(|\omega - \log\log N| \geq t(N)\sqrt{\log\log N}) \ll \frac{1}{t(N)^2}.$$

In particular, if $t(N) \to \infty$ as $N \to \infty$ then "almost all" numbers $n \leq N$ (in the sense of asymptotic density) satisfy

$$|\omega(n) - \log\log N| \leq t(N)\sqrt{\log\log N}.$$

Proofs of Corollaries 2.2 and 2.3. The Turán–Kubilius inequality immediately implies that

$$\mathbb{E}_N|\omega - \mathbb{E}_N\omega|^2 \ll \sum_{p^k \leq N} \frac{1}{p^k} \ll \log\log N \quad \text{and} \quad \mathbb{E}_N|\Omega - \mathbb{E}_N\Omega|^2 \ll \sum_{p^k \leq N} \frac{k^2}{p^k} \ll \log\log N.$$

Moreover, using our estimate for the first moment (Lemma 1.2) we have $\mathbb{E}_N\omega = \log\log N + O(1)$, and the same for Ω, so we have

$$\mathbb{E}_N(\omega - \log\log N)^2 = \mathbb{E}_N(\omega - \mathbb{E}_N\omega + O(1))^2 \ll \mathbb{E}_N(\omega - \mathbb{E}_N\omega)^2 + \mathbb{E}_N 1 \ll \log\log N,$$

and the same for Ω.

Corollary 2.3 follows immediately from Corollary 2.2 and Chebychev's inequality. □

Remark 2.4. The final statement in Corollary 2.3, that "most numbers $n \leq N$ have about $\log\log N$ prime factors", is sometimes expressed by saying that ω has *normal order* $\log\log N$. The proof we have just given was Turán's proof of this statement. The original proof, by Hardy and Ramanujan in 1917, was a 17 page long argument counting numbers with different quantities of prime factors. This shows the power of some probabilistic thinking.

3. A modern application of the Turán–Kubilius inequality

Thus far we have been concerned with fairly classical questions concerning additive functions. In this section we will show that the Turán–Kubilius inequality is relevant in some cutting edge mathematics.

Definition 3.1. A function $f : \mathbb{N} \to \mathbb{C}$ is said to be *multiplicative* if $f \not\equiv 0$ and

$$f(ab) = f(a)f(b) \quad \forall (a,b) = 1.$$

This is a natural counterpart to our definition of additive functions.

A key question in number theory is to obtain non-trivial estimates for sums of the form $\sum_{n \leq N} f(n)g(n)$, where f is multiplicative and g is some other "twist" function. For example, to prove the odd Goldbach conjecture one needs estimates for sums like

$$\sum_{n \leq N} \mu(n) e^{2\pi i \theta n},$$

where $\mu(n)$ is the Möbius function (the multiplicative function taking value -1 on all primes, and vanishing on all higher prime powers), and where θ is some real number. In a recent paper, Bourgain, Sarnak and Ziegler were interested in certain dynamical systems $(X, T : X \to X)$ called *horocycle flows*, and sought a non-trivial bound for $\sum_{n\leq N} \mu(n)g(T^n x)$, where $x \in X$ is any point in the (compact metric) space X and g is a continuous function on X.

A reason for hoping to obtain cancellation in such sums is that often the twist function is oscillatory (so there would be cancellation if the multiplicative function $f(n)$ were absent), and the structure of the twist function is not multiplicative so it should not "conspire" with $f(n)$. *One can turn this philosophy into a proof method by trying to rewrite $\sum_{n\leq N} f(n)g(n)$ using the multiplicativity of $f(n)$, producing a messier expression but one in which we see sums of $g(n)$ without a multiplier $f(n)$ (though these will be more complicated sums than simply over all $n \leq N$).*

It turns out that we can use the Turán–Kubilius inequality for the rewriting step, obtaining a non-trivial estimate given quite weak information about the twist g.

Theorem 3.2. *Let f and g be functions taking values in the complex unit disc, and suppose f is multiplicative. Let $\tau > 0$ be a small parameter, let $M_\tau \geq 1$, and suppose that for any* distinct *primes $p_1, p_2 \leq e^{1/\tau}$, and for any $M \geq M_\tau$, we have*

$$\left|\sum_{m\leq M} g(p_1 m)\overline{g(p_2 m)}\right| \leq \tau M.$$

Then

$$\left|\sum_{n\leq N} f(n)g(n)\right| \ll \frac{N}{\sqrt{\log(1/\tau)}} + \sqrt{Ne^{1/\tau}} + M_\tau e^{1/\tau}.$$

Remark 3.3. Since g takes values in the unit disc, a trivial bound for $\left|\sum_{m\leq M} g(p_1 m)\overline{g(p_2 m)}\right|$ is always M. So the hypotheses of the theorem ask for a saving of a multiplicative factor of τ, once M is large enough. Similarly, a trivial bound for $\left|\sum_{n\leq N} f(n)g(n)\right|$ is N. If τ is small then $\log(1/\tau)$ will be large, so the conclusion of the theorem gives a non-trivial bound for the twisted sum (provided N is large enough to make the second and third terms negligible).

Remark 3.4. Bourgain, Sarnak and Ziegler gave a rather complicated proof of a result like Theorem 3.2 (though with a bit stronger bound in the conclusion), which they call a *finite Vinogradov inequality*. Our proof of Theorem 3.2 is modelled on a 1986 argument of Katai, who worked in the special case of sums like $\sum_{n\leq N} f(n)e^{2\pi i\theta n}$. See Tao's blog `terrytao.wordpress.com/2011/11/21/the-bourgain-sarnak-ziegler-orthogonality-criteric` for some further discussion.

Proof of Theorem 3.2. We may assume that $N \geq e^{1/\tau}$, otherwise the conclusion is trivial because $\sqrt{Ne^{1/\tau}} > N$. For the same reason we may assume that $N/e^{1/\tau} \geq M_\tau$.

Let $\omega_\tau(n)$ denote the additive function that satisfies $\omega_\tau(p^k) = 1$ if $p \leq e^{1/\tau}$, and $\omega_\tau(p^k) = 0$ otherwise. Then by our first moment estimate (Lemma 1.2) we have

$$\mathbb{E}_N \omega_\tau = \sum_{\substack{p \leq e^{1/\tau}, \\ p^k \leq N}} \frac{1}{p^k}\left(1 - \frac{1}{p}\right) + O(\frac{\tau e^{1/\tau} \log N}{N}) = \sum_{p \leq e^{1/\tau}} \frac{1}{p} + O(1 + \frac{\tau e^{1/\tau} \log N}{N}),$$

and using Mertens' estimate (Fact 2 from Lecture Notes 0) this is all $= \log(1/\tau) + O(1 + \frac{\tau e^{1/\tau} \log N}{N})$. Since we assume $N \geq e^{1/\tau}$, the "big Oh" term is actually just $O(1)$.

Now by the Turán–Kubilius inequality (Theorem 2.1) we have

$$\begin{aligned} \sum_{n \leq N} (\omega_\tau(n) - \log(1/\tau))^2 &\ll \sum_{n \leq N} (\omega_\tau(n) - \mathbb{E}_N \omega_\tau)^2 + \sum_{n \leq N} (\mathbb{E}_N \omega_\tau - \log(1/\tau))^2 \\ &\ll N \sum_{\substack{p \leq e^{1/\tau}, \\ p^k \leq N}} \frac{1}{p^k} + N \ll N \log(1/\tau). \end{aligned}$$

Therefore we have

$$\begin{aligned} \left| \sum_{n \leq N} f(n) g(n) \right| &\leq \left| \frac{1}{\log(1/\tau)} \sum_{n \leq N} f(n) g(n) \omega_\tau(n) \right| + \left| \frac{1}{\log(1/\tau)} \sum_{n \leq N} f(n) g(n) (\omega_\tau(n) - \log(1/\tau)) \right| \\ &\leq \left| \frac{1}{\log(1/\tau)} \sum_{n \leq N} f(n) g(n) \omega_\tau(n) \right| + \frac{\sqrt{N}}{\log(1/\tau)} \sqrt{\sum_{n \leq N} (\omega_\tau(n) - \log(1/\tau))^2} \\ &\ll \left| \frac{1}{\log(1/\tau)} \sum_{n \leq N} f(n) g(n) \omega_\tau(n) \right| + \frac{N}{\sqrt{\log(1/\tau)}}, \end{aligned}$$

where the second line uses the Cauchy–Schwarz inequality and the fact that $|f(n)|, |g(n)| \leq 1$, and the third line uses our Turán–Kubilius estimate.

We have managed to insert a saving factor $1/\log(1/\tau)$ in front of our sum, at the cost of the terms $\omega_\tau(n)$, but we have used hardly any information about f or g. *Next we will show that using the factor $\omega_\tau(n)$ (which gives an extra sum, over prime divisors, to play with), we can manipulate things so that we can invoke our assumptions on g.*

Indeed, we have

$$\frac{1}{\log(1/\tau)} \sum_{n \leq N} f(n) g(n) \omega_\tau(n) = \frac{1}{\log(1/\tau)} \sum_{\substack{mp \leq N, \\ p \leq e^{1/\tau}}} f(mp) g(mp),$$

and since f is multiplicative and $|f| \leq 1$ we have $f(mp) = f(m)f(p) + O(\mathbf{1}_{p|m})$, so

$$\frac{1}{\log(1/\tau)} \sum_{n \leq N} f(n) g(n) \omega_\tau(n) = \frac{1}{\log(1/\tau)} \sum_{m \leq N} f(m) \sum_{p \leq \min\{N/m, e^{1/\tau}\}} f(p) g(mp) + O\left(\frac{N}{\log(1/\tau)}\right).$$

At this point we want to apply the Cauchy–Schwarz inequality to remove the coefficients $f(m)$. First considering the part of the sum where $m \le N/e^{1/\tau}$, we see

$$\frac{1}{\log(1/\tau)}\left|\sum_{m\le \frac{N}{e^{1/\tau}}} f(m)\sum_{p\le e^{1/\tau}} f(p)g(mp)\right| \ll \frac{1}{\log(1/\tau)}\sqrt{\frac{N}{e^{1/\tau}}}\sqrt{\sum_{m\le N/e^{1/\tau}}\left|\sum_{p\le e^{1/\tau}} f(p)g(mp)\right|^2}$$
$$= \frac{1}{\log(1/\tau)}\sqrt{\frac{N}{e^{1/\tau}}\sum_{\substack{p_1\le e^{1/\tau},\\ p_2\le e^{1/\tau}}} f(p_1)\overline{f(p_2)}\sum_{m\le \frac{N}{e^{1/\tau}}} g(mp_1)\overline{g(mp_2)}}.$$

Using our hypothesis about the sum of $g(mp_1)\overline{g(mp_2)}$ (when $p_1 \ne p_2$), together with the trivial bound $|f| \le 1$, we find everything is

$$\ll \frac{1}{\log(1/\tau)}\sqrt{\frac{N}{e^{1/\tau}}\left(\pi(e^{1/\tau})\frac{N}{e^{1/\tau}} + \sum_{p_1\ne p_2\le e^{1/\tau}} \frac{\tau N}{e^{1/\tau}}\right)} \ll \frac{N\sqrt{\tau}}{\log(1/\tau)},$$

which is more than good enough.

We will handle the part of the sum where $N/e^{1/\tau} < m \le N$ in essentially the same way, but if we directly apply the Cauchy–Schwarz inequality it will be inefficient because now the inner sums will run over $p \le N/m$, so be of rather different sizes as m varies. There is a standard and easy procedure to address this, which is to further divide the sum over m into *dyadic ranges*, as follows:

$$\frac{1}{\log(1/\tau)}\left|\sum_{\frac{N}{e^{1/\tau}}<m\le N} f(m)\sum_{p\le \frac{N}{m}} f(p)g(mp)\right| \le \frac{1}{\log(1/\tau)}\sum_{\substack{j\ge 0,\\ 2^j\le e^{1/\tau}}}\left|\sum_{\frac{2^jN}{e^{1/\tau}}<m\le \frac{2^{j+1}N}{e^{1/\tau}}} f(m)\sum_{p\le \frac{N}{m}} f(p)g(mp)\right|$$
$$\le \frac{1}{\log(1/\tau)}\sum_{\substack{j\ge 0,\\ 2^j\le e^{1/\tau}}}\sqrt{\frac{2^{j+1}N}{e^{1/\tau}}}\sqrt{\sum_{\frac{2^jN}{e^{1/\tau}}<m\le \frac{2^{j+1}N}{e^{1/\tau}}}\left|\sum_{p\le \frac{N}{m}} f(p)g(mp)\right|^2}$$
$$\ll \frac{1}{\log(1/\tau)}\sqrt{\frac{N}{e^{1/\tau}}}\sum_{\substack{j\ge 0,\\ 2^j\le e^{1/\tau}}}\sqrt{2^j}\sqrt{\sum_{p\le \frac{e^{1/\tau}}{2^j}}\sum_{\substack{\frac{2^jN}{e^{1/\tau}}<m\le \frac{2^{j+1}N}{e^{1/\tau}},\\ m\le N/p}} 1 + \sum_{p_1\ne p_2\le \frac{e^{1/\tau}}{2^j}}\left|\sum_{\substack{\frac{2^jN}{e^{1/\tau}}<m\le \frac{2^{j+1}N}{e^{1/\tau}},\\ m\le \min\{N/p_1,N/p_2\}}} g(mp_1)\overline{g(mp_2)}\right|}$$
$$\ll \frac{1}{\log(1/\tau)}\sqrt{\frac{N}{e^{1/\tau}}}\sum_{\substack{j\ge 0,\\ 2^j\le e^{1/\tau}}}\sqrt{2^j}\sqrt{\pi(e^{1/\tau}/2^j)\frac{2^jN}{e^{1/\tau}} + \sum_{p_1\ne p_2\le \frac{e^{1/\tau}}{2^j}}\frac{\tau 2^jN}{e^{1/\tau}}}.$$

This is all $\ll \frac{N}{e^{1/\tau}\log(1/\tau)}\sum_{2^j\le e^{1/\tau}} 2^j\sqrt{\pi(e^{1/\tau}/2^j) + \tau\pi(e^{1/\tau}/2^j)^2}$, and finally Chebychev's bound for the number of primes (Fact 1 from Lecture Notes 0) plus a little manipulation

shows it is $\ll N/\log(1/\tau) + N\sqrt{\tau}$, which is also more than good enough for Theorem 3.2. □

On the first problem sheet you will apply Theorem 3.2 to give a non-trivial bound for $\sum_{n\leq N}\mu(n)e^{2\pi i\theta n}$ (the case where $f(n)=\mu(n)$ and $g(n)=e^{2\pi i\theta n}$, and $\sum_{m\leq M}g(p_1m)\overline{g(p_2m)} = \sum_{m\leq M}e^{2\pi i\theta(p_1-p_2)m}$), and will see why this is useful in Goldbach-type additive problems.

4. The method of moments

We have obtained estimates for the first and second moments $\mathbb{E}_N f, \mathbb{E}_N f^2$ of additive functions, and shown their usefulness by establishing results like Corollary 2.3 (our normal order result). Now we will show that *if* we can prove that all the moments of a real-valued random variable are very close to those of the standard Gaussian (say), then our random variable must be close to the standard Gaussian in distribution. This is a powerful general method called the *method of moments*, that reappears in many places (e.g. as the *method of traces* in random matrix theory).

Proposition 4.1. *Let Z be a real-valued random variable, and let $z\in\mathbb{R}$ and $\epsilon>0$. There exist $k=k(z,\epsilon)\in\mathbb{N}$ and a small real number $\delta=\delta(z,\epsilon)>0$ such that the following is true: if we have*

$$|\mathbb{E}Z^j - m_j| \leq \delta \quad \forall j = 1,2,...,k,$$

where $m_j := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^je^{-x^2/2}dx$ are the moments of the standard normal distribution, then we have

$$|\mathbb{P}(Z\leq z) - \Phi(z)| \leq \epsilon,$$

where $\Phi(z) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-x^2/2}dx$ is the standard normal distribution function.

Remark 4.2. Notice that we have $\mathbb{P}(Z\leq z) = \mathbb{E}\mathbf{1}_{Z\leq z}$, where $\mathbf{1}$ denotes the indicator function. Therefore if it were the case that the random variable Z, and the standard normal distribution, were supported on some compact set, we could approximate the indicator by a polynomial using the Stone–Weierstrass theorem and then invoke the hypotheses about $\mathbb{E}Z^j$. Since this is not the case the proof of Proposition 4.1 will be more fiddly (to make sure the tails of the distributions do not cause problems), but the basic idea is still simply that we can approximate things by polynomials.

Lemma 4.3. *For any $N\in\mathbb{N}\cup\{0\}$, and any $x\in\mathbb{R}$, we have*

$$\left|e^{ix} - \sum_{n\leq N}\frac{(ix)^n}{n!}\right| \leq \frac{2|x|^N}{N!}.$$

Proof of Lemma 4.3. The lemma is trivial when $N = 0$, and for larger N it follows by induction on observing that

$$e^{ix} - \sum_{n\le N+1} \frac{(ix)^n}{n!} = \int_0^x ie^{iy}dy - \sum_{1\le n\le N+1} \frac{(ix)^n}{n!} = \int_0^x i\left(e^{iy} - \sum_{n\le N} \frac{(iy)^n}{n!}\right) dy.$$

□

Lemma 4.4 (Fourier Inversion Formula)**.** *Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ be a continuously differentiable function, and for each $k \in \mathbb{Z}$ define*

$$\hat{f}(k) := \int_0^1 f(x)e^{-2\pi ikx}dx.$$

Then for any $x \in [0, 1]$ and any $K \in \mathbb{N}$, we have

$$\left| f(x) - \sum_{|k|\le K} \hat{f}(k)e^{2\pi ikx} \right| \ll_f \frac{\log(2K)}{K}.$$

Proof of Lemma 4.4. This is (a special case of) a standard result on Fourier series, and so the proof is omitted. See e.g. Appendix D of Montgomery and Vaughan, *Multiplicative Number Theory.* □

We will also need to know that the moments m_j of the standard normal distribution do not grow too rapidly with j.

Lemma 4.5 (Normal Moments)**.** *If $j \in \mathbb{N}$ is odd then we have $m_j := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^j e^{-x^2/2}dx = 0$, whilst if j is even then*

$$m_j = \frac{j!}{2^{j/2}(j/2)!} \le (j/2)^{j/2}.$$

Proof of Lemma 4.5. When j is odd the result is clear because the integrand $x^j e^{-x^2/2}$ is an odd function, and when j is even it follows inductively using integration by parts. □

Proof of Proposition 4.1. Set $B = 10/\sqrt{\epsilon}$, and notice first that

$$\mathbb{P}(|Z| \ge B) \le \frac{\mathbb{E}Z^2}{B^2} \le \frac{m_2 + \delta}{B^2} \le \frac{\epsilon}{50},$$

say, provided $k \ge 2$ and $\delta \le 1$. The same (and in fact something far more precise) is true for a normal random variable. Therefore we may assume without loss of generality that $|z| \le B$, say.

Now choose any continuously differentiable functions $h^-, h^+ : \mathbb{R} \to [0, 1]$ that satisfy

$$\mathbf{1}_{[-B,z-\epsilon/10]} \le h^- \le \mathbf{1}_{[-2B,z]}, \quad \text{and} \quad \mathbf{1}_{[-B,z]} \le h^+ \le \mathbf{1}_{[-2B,z+\epsilon/10]},$$

and extend these to $6B$-periodic functions $f^-, f^+ : \mathbb{R} \to [0, 1]$ by setting $f^{\pm}(x) = h^{\pm}(\tilde{x})$, where $\tilde{x} \equiv x \bmod 6B$ is chosen so that $|\tilde{x}| \le 3B$. Since $h^{\pm}$ is continuously differentiable,

and vanishes outside the interval $[-2B, 2B]$, the $6B$-periodic extension $f^{\pm}$ will be a continuously differentiable function on $\mathbb{R}$. Moreover, since $\mathbb{P}(|Z| \geq B) \leq \epsilon/50$ we have

$$\mathbb{E}f^-(Z) - \frac{\epsilon}{50} \leq \mathbb{E}\mathbf{1}_{Z \leq z} = \mathbb{P}(Z \leq z) \leq \mathbb{E}f^+(Z) + \frac{\epsilon}{50}.$$

Therefore it will suffice to prove that

$$\left| \mathbb{E}f^-(Z) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f^-(x) e^{-x^2/2} dx \right| \leq \epsilon/2,$$

say (and the same with f^- replaced by f^+). For simplicity of writing, hereafter we shall write f rather than f^- or f^+.

Now since $x \mapsto f(6Bx)$ is a 1-periodic continuously differentiable function, the Fourier Inversion Formula (Lemma 4.4) implies that for any $K \in \mathbb{N}$ we have

$$f(6Bx) = \sum_{|k| \leq K} c_k e^{2\pi i k x} + O_{f,B}(\frac{\log(2K)}{K}),$$

where the c_k are the Fourier coefficients of $x \mapsto f(6Bx)$. Moreover, by applying Lemma 4.3 to approximate the complex exponentials, we find that for any even $N \in \mathbb{N}$ we have

$$f(6Bx) = \sum_{|k| \leq K} c_k \sum_{n \leq N} \frac{(2\pi i k x)^n}{n!} + O\left(\sum_{|k| \leq K} |c_k| \frac{(2\pi k x)^N}{N!} \right) + O_{f,B}(\frac{\log(2K)}{K}),$$

and so in particular

$$\mathbb{E}f(Z) = \mathbb{E}f(6B\frac{Z}{6B}) = \sum_{|k| \leq K} c_k \sum_{n \leq N} \frac{\mathbb{E}Z^n (2\pi i k/6B)^n}{n!} + O\left(K \frac{\mathbb{E}Z^N (2\pi K/6B)^N}{N!} \right) + O_{f,B}(\frac{\log(2K)}{K}).$$

Finally, using our estimate of the N-th normal moment (Lemma 4.5), together with the fact that $N! \geq (N/e)^N$, we see the first "big Oh" term here is

$$\ll K \left(\frac{2\pi K e}{6B} \right)^N \frac{\mathbb{E}Z^N}{N^N} \leq K \left(\frac{2\pi K e}{6B} \right)^N \frac{1}{N^{N/2}}.$$

Thus if we choose K large enough in terms of ϵ (and the fixed function f), and we choose N large enough in terms of ϵ and K, the contribution from both "big Oh" terms will be $\leq \epsilon/4$. Then if we know that $|\mathbb{E}Z^n - m_n| \leq \delta$ for all $n \leq N$, where δ is small enough in terms of K, N, ϵ, it follows as required that

$$|\mathbb{E}f^{\pm}(Z) - \mathbb{E}f^{\pm}(N(0,1))| \leq \epsilon/2.$$

□

The following immediate corollary is slightly less quantitative than Proposition 4.1, but reflects the way that the method of moments is usually applied.

Corollary 4.6 (Method of Moments)**.** *Let $(Z_n)_{n=1}^{\infty}$ be a sequence of real valued random variables, and suppose that for each fixed $j \in \mathbb{N}$ we have*

$$\mathbb{E}Z_n^j \to m_j \quad \text{as } n \to \infty,$$

where m_j are the moments of the standard normal distribution.

Then we have convergence in distribution $Z_n \xrightarrow{d} N(0,1)$ *as $n \to \infty$, in other words for all $z \in \mathbb{R}$ we have*

$$\mathbb{P}(Z_n \leq z) \to \Phi(z) \quad \text{as } n \to \infty.$$

5. The Erdős–Kac Central Limit Theorem

We remarked previously that an additive function $f(n)$ can be regarded as a sum of random variables $f_p(n)$ that are "almost independent". An obvious (to Kac!) conclusion of this line of thought is that the values of $f(n)$ should obey some kind of central limit theorem.

Theorem 5.1 (Erdős–Kac Theorem, 1939-1940)**.** *Let f be a real additive function, and suppose that $|f(p)| \leq 1$ for all primes p, that $f(p^k) = f(p)$ for all $k \in \mathbb{N}$ ("f is* strongly additive*"), and that*

$$\sum_{p \leq N} \frac{f(p)^2}{p} \to \infty \quad \text{as } N \to \infty.$$

Then under the probability measure $\mathbb{P}_N$, we have

$$\frac{f - \mathbb{E}_N f}{\sqrt{\mathbb{E}_N |f - \mathbb{E}_N f|^2}} \xrightarrow{d} N(0,1) \quad \text{as } N \to \infty.$$

Remark 5.2. The conditions that $|f(p)| \leq 1$ and that f is strongly additive can be weakened, but not removed entirely. The condition that $\sum_{p \leq N} \frac{f(p)^2}{p} \to \infty$, or in other words that the variance of f tends to infinity, is important to achieve a normal limit. All of this is analogous to the situation for sums of independent random variables, where the central limit theorem in general does require that no single variable can make too large a contribution (Lindeberg's condition) and that the variance diverges.

We saw in the proof of the Turán–Kubilius inequality (Theorem 2.1) that the "almost independence" of f_p and f_q breaks down when the product $pq > N$, and this failure becomes much more serious when trying to prove distributional theorems because one must consider many values of p simultaneously, rather than two. In the original proof of the Erdős–Kac theorem, this problem was overcome by replacing $f(n)$ by a version that is zero on all large primes (and showing this makes a negligible difference), and then using a number theoretic tool called the Brun sieve to control the failure of independence between smaller primes.

In 1969, Billingsley (building on earlier work of Delange and of Halberstam) found a much simpler proof that doesn't require the sieve. This proof, which is the one we shall broadly follow, is a well organised application of the method of moments.

Proof of Theorem 5.1. For convenience of writing, let $\phi(N) := \left(\sum_{p\leq N} \frac{f(p)^2}{p}\right)^{1/10}$. Thus $\phi(N) \to \infty$ as $N \to \infty$ by the hypotheses of the theorem, but we also have

$$\phi(N) \leq \left(\sum_{p\leq N} \frac{1}{p}\right)^{1/10} = (\log\log N + O(1))^{1/10}$$

in view of Mertens' estimate (Fact 2 from Chapter 0), so $\phi(N)$ does not tend to infinity very quickly. In particular, we have $N^{1/\phi(N)} \gg e^{(\log N)/(\log\log N)^{1/10}} \to \infty$. Let $g(n) = g_N(n)$ denote the strongly additive function that agrees with f on all prime powers p^k with $p \leq N^{1/\phi(N)}$, but that equals zero on all larger primes.

To prove the Erdős–Kac theorem, we will first use the method of moments to show that the truncated function $g_N(n)$ satisfies

$$\frac{g_N - \mathbb{E}_N g_N}{\sqrt{\mathbb{E}_N |g_N - \mathbb{E}_N g_N|^2}} \xrightarrow{d} N(0,1) \quad \text{as } N \to \infty,$$

and then will show that the difference between f and g_N is sufficiently small to imply the analogous result for f.

Lemma 5.3 (Truncated Moments)**.** *In the above situation, for any large N and even $j \in \mathbb{N}$ we have*

$$\mathbb{E}_N(g_N(n) - \mathbb{E}_N g_N)^j = m_j \left(\sum_{p\leq N^{1/\phi(N)}} \frac{f(p)^2}{p}\left(1-\frac{1}{p}\right)\right)^{j/2} (1+O_j(\frac{1}{\sum_{p\leq N^{1/\phi(N)}} \frac{f(p)^2}{p}}))+O(\frac{jN^{2j/\phi(N)}}{N})$$

(where $m_j = \frac{j!}{2^{j/2}(j/2)!}$ denotes the standard normal moment), whilst for any odd $j \in \mathbb{N}$ we have

$$\mathbb{E}_N(g_N(n) - \mathbb{E}_N g_N)^j \ll_j \left(\sum_{p\leq N^{1/\phi(N)}} \frac{f(p)^2}{p}\left(1-\frac{1}{p}\right)\right)^{(j-1)/2} + \frac{N^{2j/\phi(N)}}{N}.$$

Proof of Lemma 5.3. Notice that since g_N is strongly additive, we have

$$g_N(n) - \mathbb{E}_N g_N = \sum_{\substack{p|n,\\ p\leq N^{1/\phi(N)}}} f(p) - \mathbb{E}_N g_N = \sum_{p\leq N^{1/\phi(N)}} f(p)(\mathbf{1}_{p|n} - \mathbb{E}_N \mathbf{1}_{p|n}),$$

where $\mathbf{1}$ denotes the indicator function. It is this representation (that is extremely natural when thinking of sums of random variables centred to have mean zero) that is key to organising the moment computation efficiently.

Indeed, if we now expand out we find that

$$\mathbb{E}_N\left(\sum_{p\leq N^{1/\phi(N)}} f(p)(\mathbf{1}_{p|n}-\mathbb{E}_N\mathbf{1}_{p|n})\right)^j = \sum_{p_1,\dots,p_j\leq N^{1/\phi(N)}} \mathbb{E}_N\prod_{i=1}^j f(p_i)(\mathbf{1}_{p_i|n}-\mathbb{E}_N\mathbf{1}_{p_i|n}).$$

As usual we have that $\mathbb{E}_N\mathbf{1}_{p_i|n}=1/p_i+O(1/N)$, and therefore (using also that $|f(p_i)(\mathbf{1}_{p_i|n}-\mathbb{E}_N\mathbf{1}_{p_i|n})|\leq 1$) we have

$$\begin{aligned}\sum_{p_1,\dots,p_j\leq N^{1/\phi(N)}} \mathbb{E}_N\prod_{i=1}^j f(p_i)(\mathbf{1}_{p_i|n}-\mathbb{E}_N\mathbf{1}_{p_i|n}) &= \sum_{p_1,\dots,p_j\leq N^{1/\phi(N)}} \mathbb{E}_N\prod_{i=1}^j f(p_i)(\mathbf{1}_{p_i|n}-1/p_i)\\ &\quad +O\left(\sum_{p_1,\dots,p_j\leq N^{1/\phi(N)}}\frac{j}{N}\right).\end{aligned}$$

Here the "big Oh" term is acceptably small. Further, if we let $R(=q_1...q_w)$ denote the product of all the distinct primes amongst $p_1,...,p_j$ (so that $p_1...p_j=q_1^{a_1}...q_w^{a_w}$, say), and if we let X_p denote independent random variables taking value 1 with probability $1/p$, and taking value 0 otherwise, then

$$\sum_{n\leq R}\prod_{i=1}^j(\mathbf{1}_{p_i|n}-1/p_i) = R\mathbb{E}\prod_{i=1}^j(X_{p_i}-1/p_i) = R\prod_{i=1}^w\mathbb{E}(X_{q_i}-1/q_i)^{a_i},$$

since there are no "big Oh" error terms when summing over a complete set of R residues. We can then conclude that

$$\begin{aligned}\mathbb{E}_N\prod_{i=1}^j f(p_i)(\mathbf{1}_{p_i|n}-\mathbb{E}_N\mathbf{1}_{p_i|n}) &= \left(\prod_{i=1}^j f(p_i)\right)\frac{1}{N}\left(\lfloor\frac{N}{R}\rfloor R\prod_{i=1}^w\mathbb{E}(X_{q_i}-1/q_i)^{a_i}+O(R)\right)\\ &= \prod_{i=1}^w f(q_i)^{a_i}\mathbb{E}(X_{q_i}-1/q_i)^{a_i}+O(\frac{R}{N}),\end{aligned}$$

and the overall contribution from the "big Oh" term is $\ll\sum_{p_1,\dots,p_j\leq N^{1/\phi(N)}}\frac{p_1...p_j}{N}\leq\frac{N^{2j/\phi(N)}}{N}$. *Now we have reduced to dealing with truly independent random variables.*

If j is even, then one collection of terms in $\sum_{p_1,\dots,p_j\leq N^{1/\phi(N)}}$ is those where we have $j/2$ pairs of distinct primes. The contribution from these terms is

$$\sum_{\substack{q_1,\dots,q_{j/2}\leq N^{1/\phi(N)},\\ \text{distinct primes}}}\frac{j!}{2^{j/2}(j/2)!}\prod_{i=1}^{j/2}f(q_i)^2\mathbb{E}(X_{q_i}-1/q_i)^2 = \sum_{\substack{q_1,\dots,q_{j/2}\leq N^{1/\phi(N)},\\ \text{distinct primes}}} m_j\prod_{i=1}^{j/2}\frac{f(q_i)^2}{q_i}\left(1-\frac{1}{q_i}\right)$$

$$\begin{aligned}&= m_j\Big(\sum_{p\leq N^{1/\phi(N)}}\frac{f(p)^2}{p}\left(1-\frac{1}{p}\right)\Big)^{j/2}\\ &\quad +O\Big(j^2m_j\sum_{p\leq N^{1/\phi(N)}}\frac{f(p)^4}{p^2}\left(1-\frac{1}{p}\right)^2\Big(\sum_{p\leq N^{1/\phi(N)}}\frac{f(p)^2}{p}\left(1-\frac{1}{p}\right)\Big)^{(j/2)-2}\Big),\end{aligned}$$

where the first term is obtained by ignoring the distinctness condition on the q_i, and the "big Oh" term corrects for this overcount. This all gives the main term in the statement of the lemma.

To bound the contribution from all the other terms in $\sum_{p_1,\dots,p_j \le N^{1/\phi(N)}}$, where the primes p_i do *not* all just match in pairs, let us note that $\mathbb{E}(X_q - 1/q) = 0$ (so every prime must appear with multiplicity at least two to give a non-zero contribution), and for any $a \ge 2$ we have $\mathbb{E}(X_q - 1/q)^a \le \mathbb{E}(X_q - 1/q)^2 = \frac{1}{q}(1 - \frac{1}{q})$. So this contribution is

$$\sum_{w<j/2} \sum_{\substack{p_1,\dots,p_j \le N^{1/\phi(N)}, \\ w \text{ primes are distinct}, \\ p_1\dots p_j = q_1^{a_1}\dots q_w^{a_w},\ a_i \ge 2}} \prod_{i=1}^{w} f(q_i)^{a_i} \mathbb{E}(X_{q_i} - 1/q_i)^{a_i} \quad \le \quad \sum_{w<j/2} \sum_{\substack{p_1,\dots,p_j \le N^{1/\phi(N)}, \\ w \text{ primes are distinct}, \\ p_1\dots p_j = q_1^{a_1}\dots q_w^{a_w},\ a_i \ge 2}} \prod_{i=1}^{w} \frac{f(q_i)^2}{q_i}\left(1 - \frac{1}{q_i}\right)$$

$$\ll_j \sum_{w<j/2} \Big(\sum_{p \le N^{1/\phi(N)}} \frac{f(p)^2}{p}\left(1 - \frac{1}{p}\right)\Big)^w,$$

where any combinatorial factors (counting the number of tuples of $p_1, \dots, p_j$ for which $p_1 \dots p_j = q_1^{a_1} \dots q_w^{a_w}$, for given primes q_i) are absorbed in the implicit constant $\ll_j$. If j is even then the largest term here has $w = (j/2) - 1$, whereas if j is odd the largest term has $w = (j-1)/2$, which in any case is acceptable to finish the proof of Lemma 5.3. $\square$

The special case of the Truncated Moments lemma where $j = 2$ implies (since $\sum_{p \le N^{1/\phi(N)}} \frac{f(p)^2}{p}\left(1 - \frac{1}{p}\right) \to \infty$ and $\phi(N) \to \infty$) that

$$\mathbb{E}_N(g(n) - \mathbb{E}_N g_N)^2 \sim m_2 \sum_{p \le N^{1/\phi(N)}} \frac{f(p)^2}{p}\left(1 - \frac{1}{p}\right) = \sum_{p \le N^{1/\phi(N)}} \frac{f(p)^2}{p}\left(1 - \frac{1}{p}\right) \quad \text{as } N \to \infty,$$

and on feeding this back into the lemma we obtain that for any fixed $j \in \mathbb{N}$,

$$\mathbb{E}_N \left(\frac{g(n) - \mathbb{E}_N g_N}{\sqrt{\mathbb{E}_N (g(n) - \mathbb{E}_N g_N)^2}} \right)^j \to m_j \quad \text{as } N \to \infty.$$

Therefore the Method of Moments (Corollary 4.6) indeed implies that $\frac{g_N - \mathbb{E}_N g_N}{\sqrt{\mathbb{E}_N (g_N - \mathbb{E}_N g_N)^2}} \xrightarrow{d} N(0,1)$ as $N \to \infty$.

Further, for any $n \le N$ we have $f(n) - g_N(n) = \sum_{p | n, p > N^{1/\phi(N)}} f(p) = O(\phi(N))$, whilst by the above calculations and the definition of $\phi(N)$ we see $\mathbb{E}_N(g(n) - \mathbb{E}_N g_N)^2$ is

$$\gg \sum_{p \le N^{1/\phi(N)}} \frac{f(p)^2}{p} = \phi(N)^{10} - \sum_{N^{1/\phi(N)} < p \le N} \frac{f(p)^2}{p} \ge \phi(N)^{10} - \sum_{N^{1/\phi(N)} < p \le N} \frac{1}{p} \gg \phi(N)^{10}.$$

(Here we used that $\sum_{N^{1/\phi(N)} < p \le N} \frac{1}{p} = \log \phi(N) + O(1)$.) Thus

$$\frac{f - \mathbb{E}_N f}{\sqrt{\mathbb{E}_N |f - \mathbb{E}_N f|^2}} = \frac{g_N - \mathbb{E}_N g_N + O(\phi(N))}{\sqrt{\mathbb{E}_N (g_N - \mathbb{E}_N g_N + O(\phi(N)))^2}} = \frac{g_N - \mathbb{E}_N g_N + O(\phi(N))}{\sqrt{\mathbb{E}_N (g_N - \mathbb{E}_N g_N)^2 + O(\phi(N)^6)}},$$

and since $\mathbb{E}_N(g_N - \mathbb{E}_N g_N)^2 + O(\phi(N)^6) = (1 + O(1/\phi(N)^4))\mathbb{E}_N(g_N - \mathbb{E}_N g_N)^2$, the convergence in distribution $\frac{f-\mathbb{E}_N f}{\sqrt{\mathbb{E}_N|f-\mathbb{E}_N f|^2}} \xrightarrow{d} N(0,1)$ follows. □

It is worth noting that the above proof was completed with quite a lot of room to spare— our choice of the truncation $N^{1/\phi(N)}$ in the definition of g_N was not especially delicate (many other choices would also work), and our handling of the "big Oh" terms when calculating $\mathbb{E}_N \prod_{i=1}^{j} f(p_i)(\mathbf{1}_{p_i|n} - \mathbb{E}_N \mathbf{1}_{p_i|n})$ gave much sharper estimates than we really needed. We will comment on this again in the final section of this chapter.

Returning to our favourite special case of the number of prime factors function ω (where $\omega(p^k) = 1$ for all primes p and all $k \in \mathbb{N}$), the Erdős–Kac theorem takes the following form.

Corollary 5.4. *Under the probability measure $\mathbb{P}_N$, we have*

$$\frac{\omega - \log\log N}{\sqrt{\log\log N}} \xrightarrow{d} N(0,1) \quad \text{as } N \to \infty.$$

Proof of Corollary 5.4. Lemma 1.2 implies that $\mathbb{E}_N\omega = \log\log N + O(1)$, and the calculations at the end of the proof of the Erdős–Kac theorem imply that the variance $\mathbb{E}_N(\omega - \mathbb{E}_N\omega)^2 \sim \sum_{p\le N} \frac{1}{p}\left(1 - \frac{1}{p}\right) = \log\log N + O(1)$. Corollary 5.4 follows by combining these facts with the general statement of the Erdős–Kac theorem (Theorem 5.1). □

Corollary 5.4 hopefully seems like an appealing result in its own right, but we shall also record a quick consequence that you will explore further on the first problem sheet.

Let $d(n) = \sum_{d|n} 1$ denote the divisor function.

Corollary 5.5. *For any large N, at least $N/2 + o(N)$ integers $n \le N$ satisfy*

$$d(n) \ge (\log N)^{\log 2}.$$

Proof of Corollary 5.5. Note that $d(n) = \prod_{p^k||n}(k+1) \ge 2^{\omega(n)}$, and so

$$\frac{1}{N}\#\{n \le N : d(n) \ge (\log N)^{\log 2}\} \ge \frac{1}{N}\#\{n \le N : \omega(n) \ge \log\log N\} \to \frac{1}{2} \quad \text{as } N \to \infty,$$

by the Erdős–Kac theorem. □

It is worth noting that if we try to study the distribution of $d(n)$ by directly estimating its first moment, say, then we obtain

$$\mathbb{E}_N d(n) = \mathbb{E}_N \sum_{d\le N} \mathbf{1}_{d|n} = \sum_{d\le N} \mathbb{P}_N(\mathbf{1}_{d|n}) = \sum_{d\le N}(1/d + O(1/N)) = \log N + O(1).$$

Thus the first moment is of a substantially different size than the range of values $(\log N)^{\log 2}$ considered in the corollary.

6. Closing thoughts about additive functions

The results we obtained in this chapter develop an analogue of very classical probability theory (moment estimates, central limit theorem) for additive functions. We end by mentioning two directions in which these investigations can be extended further.

Question 6.1. *What is the rate of convergence in the Erdős–Kac theorem? For example, what is the smallest function $E(N)$ for which it is true that*

$$\left| \mathbb{P}_N \left(\frac{\omega - \log\log N}{\sqrt{\log\log N}} \leq z \right) - \Phi(z) \right| \leq E(N) \quad \forall z \in \mathbb{R} \quad ?$$

Corollary 5.4 (applied at a suitable "net" of points z) implies that we can take $E(N) = o(1)$, but it gives no stronger information. (This is connected to the relatively "soft" nature of the proof, which didn't require very strong or uniform estimates.) In the other direction, let $y_N := \frac{\lfloor \log\log N \rfloor - \log\log N}{\sqrt{\log\log N}}$ and let $z_N := \frac{\lfloor \log\log N \rfloor + 1/2 - \log\log N}{\sqrt{\log\log N}}$, and note that if the above inequality is satisfied then we must have

$$\left| \mathbb{P}_N \left(\frac{\omega - \log\log N}{\sqrt{\log\log N}} \leq z_N \right) - \mathbb{P}_N \left(\frac{\omega - \log\log N}{\sqrt{\log\log N}} \leq y_N \right) - (\Phi(z_N) - \Phi(y_N)) \right| \leq 2E(N).$$

Since ω takes integer values it is clear that $\mathbb{P}_N \left(\frac{\omega - \log\log N}{\sqrt{\log\log N}} \leq z_N \right) - \mathbb{P}_N \left(\frac{\omega - \log\log N}{\sqrt{\log\log N}} \leq y_N \right) = 0$, whereas

$$\Phi(z_N) - \Phi(y_N) = \frac{1}{\sqrt{2\pi}} \int_{y_N}^{z_N} e^{-t^2/2} dt \gg (z_N - y_N) \gg \frac{1}{\sqrt{\log\log N}},$$

so we must have $E(N) \gg \frac{1}{\sqrt{\log\log N}}$. This is the usual "continuity correction" error when approximating discrete random variables by continuous ones.

It was conjectured by LeVeque (by analogy with the Berry–Esseen theorem on the rate of convergence in the central limit theorem) that actually one can take $E(N) = \frac{C}{\sqrt{\log\log N}}$, and this was proved by Rényi and Turán in 1958, but their proof involves rather sophisticated complex analysis. In my opinion, it remains an open problem to find a really nice probabilistic proof of the Rényi–Turán theorem.

Question 6.2. *We showed that ω has an approximately Gaussian distribution, but this is true of many different random variables. What more can we say specifically about the probabilistic structure of ω (or other additive functions)?*

Note that $\omega(n)$ takes integer values; is a sum of many almost independent random variables that occur with small probability; and has mean and variance approximately the same ($= \log\log N + O(1)$). This strongly suggests that $\omega(n)$ might really have an approximately Poisson distribution (which is also approximately normal when the variance tends to infinity). In fact it is a classical result of Landau that for any *fixed*

$k \in \mathbb{N}$, we have the Poisson-type probabilities

$$\mathbb{P}_N(\omega = k) = (1 + o(1))\frac{1}{\log N}\frac{(\log\log N)^{k-1}}{(k-1)!} \quad \text{as } N \to \infty.$$

This can be developed in various ways, for example obtaining results that are uniform as $k \to \infty$ at a certain rate (done by Sathe and Selberg), or proving a *total variation Poisson approximation* for the distribution of ω.

Jesus College, Cambridge, CB5 8BL
E-mail address: A.J.Harper@dpmms.cam.ac.uk