THE ZETA FUNCTION AND PRIME NUMBERS

H. L. Montgomery

1. Zeros of $\sum_{n\le N} n^{-s}$.
2. Zeros on the critical line.
3. Sums of independent random variables.
4. Vaughan's version of Vinogradov's method.

PREFACE: In these lectures we present unrelated results. In the first lecture we discuss work of Turán concerning the Riemann Hypothesis. In the second lecture we give a brief account of Levinson's work concerning zeros of the zeta function on the critical line. If the Riemann Hypothesis is true then the error term in the prime number theorem has a limiting distribution (after normalization and a change of variables) which can be related to sums of random variables; we investigate this in Lecture 3. Finally in Lecture 4 we present Vaughan's improved version of a method of Vinogradov for estimating sums of the sort $\sum_{p\le x} f(p)$.

1. ZEROS OF $\sum_{n\leq N} n^{-s}$.

Turán proved that if for all large N $\zeta_N(s) = \sum_{n\leq N} n^{-s}$ does not vanish in the half plane $\sigma > 1$ then the Riemann Hypothesis is true. We shall show presently that this hypothesis is false, but as the argument may have some use in a modified form, we first give Turán's proof.

Suppose that $\zeta_N(s) \neq 0$ for $\sigma > 1$. Then by Bohr's theory of values of absolutely convergent Dirichlet series (as found, for example, in the last chapter of Apostol's "Modular Functions and Dirichlet Series in Number Theory"), it follows that if f is totally multiplicative and unimodular, then the Dirichlet polynomial $\sum_{n\leq N} \frac{f(n)}{n^s}$ does not vanish in the half-plane $\sigma > 1$. We take the particular case $f(n) = \lambda(n)$, where $\lambda(n)$ is Liouville's lambda function, $\lambda(n) = (-1)^{\Omega(n)}$. (Here $\Omega(n)$ is the total number of prime factors of n, counting multiplicity.) From the fact that $\sum_{n\leq N} \frac{\lambda(n)}{n^s}$ does not vanish in the half-plane $\sigma > 1$ it follows in particular that it does not vanish on the ray $\sigma > 1$. That is, $\sum_{n\leq N} \frac{\lambda(n)}{n^\sigma} \neq 0$ for $\sigma > 1$. But this is a continuous function which tends to 1 as σ tends to ∞; hence $\sum_{n\leq N} \frac{\lambda(n)}{n^\sigma} > 0$ for $\sigma > 1$, and in particular $\sum_{n\leq N} \frac{\lambda(n)}{n} \geq 0$ for all large N. We put now $L(u) = \sum_{n\leq u} \frac{\lambda(n)}{n}$. This is a coefficient sum of a Dirichlet series, which we recognize by its Euler product:

$$\sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} = \prod_p \left(1 - \frac{1}{p^s} + \frac{1}{p^{2s}} - \dots\right) = \prod_p \left(1 + \frac{1}{p^s}\right)^{-1} = \frac{\zeta(2s)}{\zeta(s)} .$$

We now relate $\frac{\zeta(2s)}{\zeta(s)}$ to the coefficient sum $L(u)$. In general, if we have a Dirichlet series $A(s) = \sum_{n=1}^{\infty} a_n n^{-s}$, then

(1) $$A(s) = s\int_1^{\infty} \left(\sum_{n \leq N} a_n\right) \frac{du}{u^{s+1}}$$

To see this, write $n^{-s} = s\int_n^{\infty} \frac{du}{u^{s+1}}$. Then

$$A(s) = s \sum_{n=1}^{\infty} a_n \int_n^{\infty} \frac{du}{u^{s+1}} = s \int_1^{\infty} \left(\sum_{n \leq N} a_n\right) \frac{du}{u^{s+1}} .$$

We have not said when (1) is valid. The argument just given shows that (1) holds when $\sigma > \max(0, \sigma_a)$. By arguing more carefully it can be shown that (1) holds when $\sigma > \max(0, \sigma_c)$. In the case at hand we have

(2) $$\frac{\zeta(2s)}{\zeta(s)} = (s-1)\int_1^{\infty} \frac{L(u)}{u^s}\, du .$$

A general theorem of Landau asserts that if a general Dirichlet series $A(s) = \sum a_n e^{-\lambda_n s}$ has non-negative coefficients and abscissa of convergence σ_c, then the point σ_c is a singular point of $A(s)$. This theorem applies also to the integral in (2), and hence σ_c is a singularity of the meromorphic function

(3) $$\int_1^{\infty} \frac{L(u)}{n^s}\, du = \frac{\zeta(2s)}{(s-1)\zeta(s)} .$$

But it is easy to check that the only real pole of $\frac{\zeta(2s)}{(s-1)\zeta(s)}$ is at $s = \frac{1}{2}$. Hence the integral (3) converges for $\sigma > \frac{1}{2}$, so that $\frac{\zeta(2s)}{(s-1)\zeta(s)}$ is regular in this half-plane. But if $\rho = \beta + i\gamma$ were a zero of $\zeta(s)$ with $\beta > \frac{1}{2}$ then $\frac{\zeta(2s)}{(s-1)\zeta(s)}$ would have a pole at ρ; hence no such zero exists, and the Riemann Hypothesis is true.

Turán also found a bound for the zeros of $\zeta_N(s)$: If N is sufficiently large then $\zeta_N(s) \neq 0$ for

$$\sigma > 1 + \frac{2 \log \log N}{\log N} \qquad (4)$$

To see this, note that $\zeta_N(s) = \zeta(s) - \sum_{n>N} n^{-s}$, and hence

$$|\zeta_N(s)| \geq |\zeta(s)| - \sum_{n>N} n^{-\sigma}$$

But for $\sigma > 1$,

$$\inf_t |\zeta(\sigma+it)| = \inf_t \prod_p |-p^{-\sigma-it}|^{-1} = \prod_p (1+p^{-\sigma})^{-1} = \frac{\zeta(2\sigma)}{\zeta(\sigma)},$$

and this is $> \frac{(\sigma-1)}{\sigma}$. On the other hand

$$\sum_{n>N} n^{-\sigma} \leq \int_N^\infty u^{-\sigma} du = \frac{N^{1-\sigma}}{\sigma-1},$$

so that

$$|\zeta_N(s)| > \frac{(\sigma-1)}{\sigma} - \frac{N^{1-\sigma}}{\sigma-1} > 0$$

if $\sigma \geq 1 + \frac{2 \log \log N}{\log N}$ and $N > N_0$.

I have shown (in a paper to appear in a volume dedicated to Turán's memory) that Turán's hypothesis is badly false, and in fact that the bound (4) is close to the truth. (Turán never conjectured that his hypothesis is true, he merely showed that it implies the Riemann Hypothesis.) More precisely, if $c < \frac{4}{\pi} - 1$ and $N > N_0(c)$ then $\zeta_N(s)$ has zeros in the half-plane

(5) $$\sigma > 1 + \frac{c \log \log N}{\log N} .$$

By a different method, using Halász's method for estimating sums of multiplicative functions, I have also shown that if $c > \frac{4}{\pi} - 1$ and $N > N_0(c)$ then $\zeta_N(s) \neq 0$ for s in the half-plane (5).

We now give a simple proof of a weaker result, namely that there is an absolute constant $c > 0$ such that for all large N, $\zeta_N(s)$ has a zero in the half-plane

(6) $$\sigma > 1 + \frac{c}{\log N} .$$

Let a_n and b_n be unimodular totally multiplicative functions. By Bohr's theory it suffices to construct b_n so that $\sum_{n \leq N} b_n n^{-s}$ has a zero in the half-plane (6). We start with a sum $\sum_{n \leq N} a_n n^{-s}$ which has a zero near 1, and seek to move it to the right by altering the definition of the a_n. Such an alteration on several primes means that we must alter a_n on all multiples of these primes, which leads to complications. However, if we alter a_n only on primes $p \in (\frac{N}{2}, N]$ then no other terms $n \leq N$ are affected. Accordingly we take

$$b_p = \begin{cases} e(\theta_p) & \text{if } p \in (\frac{N}{2}, N], \\ a_p & \text{otherwise;} \end{cases}$$

here the $e(\theta_p)$ are selected later. Then

$$(7) \qquad \sum_{n \leq N} \frac{b_n}{n^s} = \sum_{n \leq N} \frac{a_n}{n^s} + \sum_{\frac{N}{2} < p \leq N} \frac{1}{p^s}(e(\theta_p) - a_p).$$

We try first $a_n = \lambda(n)$. By the prime number theorem,

$$(8) \qquad \sum_{n \leq N} \frac{\lambda(n)}{n^s} = \frac{\zeta(2s)}{\zeta(s)} + O(N^{1-\sigma}\exp(-\sqrt{\log N}))$$

for $\sigma \geq 1$, $|t| \leq \exp(\sqrt{\log N})$. If we take $s = \sigma$ in (7) then the sum over p takes on all values in the disc $|z-r| \leq r$ as the $e(\theta_p)$ vary, where

$$r = \sum_{\frac{N}{2} < p \leq N} \frac{1}{p^\sigma} \approx \frac{N^{1-\sigma}}{\log N} \quad .$$

This does not work, for in (7) both terms have positive real part. The further idea needed here is that by taking a slightly more complicated choice of a_n we can essentially change the sign of the main term on the right in (8). Take

$$a_p = \begin{cases} i & \text{if } p \le P , \\ -1 & \text{if } p > P , \end{cases}$$

and let P be an absolute constant determined by the property that P is the least number for which

$$\arg \prod_{p \le P} \left(1 - \frac{i}{p}\right)^{-1} > \pi .$$

Then

$$\sum_{n \le N} \frac{a_n}{n^s} = \frac{\zeta(2s)}{\zeta(s)} \prod_{p \le P} \left(1 - \frac{i}{p^s}\right)^{-1} \left(1 + \frac{1}{p^s}\right) + O(N^{1-\sigma}\exp(-\sqrt{\log N}))$$

for $\sigma \ge 1$, $|t| \le \exp(\sqrt{\log N})$. Hence in particular

$$\sum_{n \le N} \frac{a_n}{n^\sigma} = \frac{\zeta(2\sigma)}{\zeta(\sigma)} A(\sigma) + O\left(\frac{1}{(\log N)^2}\right)$$

for $\sigma \ge 1$, where $A(\sigma)$ is continuous, $A'(\sigma) \ll 1$, and $\arg A(1) = \pi + \delta$ with $0 < \delta < \frac{1}{100}$. Now we see that if $\sigma = 1 + \frac{c}{\log N}$, c small and positive, then a suitable choice of $e(\theta_p)$ in (7) gives $\sum_{n \le N} \frac{b_n}{n^\sigma} = 0$.

Note that our argument shows that the real parts of the zeros of $\sum_{n \le N} n^{-s}$ are dense in the interval $[1, 1 + \frac{c}{\log N}]$. Reynolds Monach has used a similar argument to show that $\zeta_N(s)$ has a zero in $\sigma > 1$ for all $N > 30$.

2. ZEROS ON THE CRITICAL LINE

Let $N(t)$ denote the number of zeros of $\zeta(s)$ for which $0 < \beta < 1$, $0 < \gamma \leq t$, let $N_0(T)$ denote the number of zeros for which $\beta = \frac{1}{2}$, $0 < \gamma \leq T$, and let $N_1(T)$ denote the number of simple zeros such that $\beta = \frac{1}{2}$, $0 < \gamma \leq T$. Then $N(T) \sim \frac{T}{2\pi} \log T$ as $T \to \infty$. It was proved by Hardy in 1914 that $N_0(T) \to \infty$. Later $N_0(T) > cT$ was obtained in several ways; the proof of Siegel (Collected Works) is notable. In 1942 Selberg proved that that $N_0 > cN(T)$, so that a positive proportion of the zeros are on the critical line. In 1974, Levinson (Advances Math. 13 (1974), 383-436) devised a new method which gives $c = 0.34$. Later Selberg (unpublished) and Heath-Brown (Bull. London Math. Soc. independently observed that with a slight modification the same lower bound applies to $N_1(T)$. We give an account of Levinson's method.

We begin by proving a theorem of Levinson and Montgomery (Acta Math 133 (1974), 49-65).

THEOREM 1. Let $N_L(T)$ and $N_L'(T)$ denote the number of zeros of $\zeta(s)$ and of $\zeta'(s)$, respectively, in the rectangle $-1 < \sigma < \frac{1}{2}$, $0 < t \leq T$. Then $N_L'(T) = N_L(T) + O(\log T)$.

It was proved by Speiser (Math. Ann. 110 (1934), 514-521) that the Riemann Hypothesis is equivalent to the assertion that $\zeta'(s) \neq 0$ for $0 < \sigma < \frac{1}{2}$

Proof: We examine the change in argument of $\frac{\zeta'}{\zeta}(s)$ along the path from $\frac{1}{2} + iA$ to $\frac{1}{2} + iT$, to $-1 + iT$ to $-1 + iA$. To avoid passing through a zero of $\zeta(s)$ on the critical line we take a semicircular detour to the left around the zero. Let $h(s) = \Gamma(\frac{s}{2})\pi^{-s/2}$. Then $\zeta(s)h(s)$ is real when $\sigma = \frac{1}{2}$; consequently $\frac{\zeta'}{\zeta}(\frac{1}{2}+it) + \frac{h'}{h}(\frac{1}{2}+it)$ is purely imaginary. But $\frac{h'}{h}(s) = \frac{1}{2}\log s + O(1)$ so that $\frac{\zeta'}{\zeta}(\frac{1}{2}+it) < 0$ along our path from $\frac{1}{2} + iA$ to $\frac{1}{2} + iT$. To treat the other vertical range we take logarithmic derivatives of the functional equation, $\zeta(s)h(s) = \zeta(1-s)h(1-s)$, to get

$$\frac{\zeta'}{\zeta}(s) = -\frac{\zeta'}{\zeta}(1-s) - \frac{h'}{h}(s) - \frac{h'}{h}(1-s) .$$

Taking $s = -1 + it$, we see that $\frac{\zeta'}{\zeta}(-1+it) = -\log t + O(1)$. Hence $\mathrm{Re}\,\frac{\zeta'}{\zeta}(-1+it) < 0$ on this path also. The change of argument of $\frac{\zeta'}{\zeta}(s)$ on the horizontal segments is $O(\log T)$, as we see from Lemma 9.4 of Titchmarsh's Theory of the Riemann zeta function. Hence the result.

We now outline the proof of

THEOREM 2. For all large T, $N_1(T) > (0.34)\frac{T}{2\pi}\log T$.

Let $N_0'(T)$ denote the number of zeros of $\zeta'(s)$ with real part $\frac{1}{2}$ and imaginary part between 0 and T . By the proof of Theorem 1 we see that $\zeta'(\frac{1}{2}+it) = 0$ can happen only when $\zeta(\frac{1}{2}+it) = 0$. Hence if ρ is a zero of $\zeta(s)$ of multiplicity m_ρ then it is a zero of $\zeta'(s)$ of multiplicity $m_\rho - 1$, so that $N_0'(T) = \sum_{\substack{0<\gamma\le T \\ \text{distinct } \gamma \\ \beta = \frac{1}{2}}} (m_\rho - 1)$.

Let $N_G(T) = N_L'(T) + N_0'(T)$, so that $N_G(T)$ is the number of zeros of $\zeta'(s)$ in the closed rectangle $-1 \le \sigma \le \frac{1}{2}$, $A \le t \le T$. We write the functional equation as $\zeta(1-s) = \frac{h(s)}{h(1-s)}\zeta(s)$, and differentiate to obtain the identity

$$\zeta'(1-s) = -\frac{h(s)}{h(1-s)}\left(\left(\frac{h'}{h}(s)+\frac{h'}{h}(1-s)\right)\zeta(s)+\zeta'(s)\right)$$

$$= -\frac{h(s)}{h(1-s)}\, G(s),$$

say. The factor $h(s)/h(1-s)$ is regular and non-vanishing, so that $N_G(T)$ is the number of zeros of G in the rectangle $\frac{1}{2} \le \sigma \le 2$, $A \le t \le T$. Since $N(T) = N_0(T) + 2N_L(T)$, it follows from Theorem 1 that

$$N(T) - 2N_G(T) + O(\log T) = \sum_{\substack{0<\gamma\le T \\ \gamma \text{ distinct} \\ \beta = \frac{1}{2}}} (2-m_\rho) \le N_1(T) .$$

Thus it suffices to show that $N_G(T) < \frac{0.33}{2\pi} T \log T$ for all large T. To this end we employ a method which is standard in the theory of the zeta function. First we have

Littlewood's Lemma. If $F(s)$ is regular in the rectangle $a \leq \sigma \leq 2$, $T \leq t \leq T+U$, and if $F(s)$ has zeros $z_1, \ldots, z_k$ in this rectangle, then

$$(1) \quad \sum_{k=1}^{K} (\operatorname{Re} z_k - a) = \frac{1}{2\pi} \int_T^{T+U} \log|F(a+it)|\,dt$$

$$- \frac{1}{2\pi} \int_T^{T+U} \log|F(2+it)|\,dt$$

$$+ \frac{1}{2\pi} \int_a^2 \arg F(\sigma+i(T+U))\,d\sigma$$

$$- \frac{1}{2\pi} \int_a^3 \arg F(\sigma+iT)\,d\sigma \quad .$$

This is an analogue of Jensen's formula, and is proved in the same way. Since we want an upper bound for $N_G(T+U) - N_G(T)$ we must take $a < \frac{1}{2}$; in fact we take $a = \frac{1}{2} - \frac{R}{\log T}$, where R is a parameter to be chosen later. Hence

$$N_G(T+U) - N_G(T) \leq \frac{\log T}{R} \sum_{k=1}^{K} (\operatorname{Re} z_k - a)$$

We do not take $F = G$, but rather $F = G\psi$, where $\psi(s)$ is chosen to smooth out the peaks of G, so that our estimate for the first term on the right in (1) is improved. This may introduce further zeros, i.e. zeros of ψ, but this only makes the sum (1) larger. Levinson chose the mollifier

$$\psi(s) = \sum_{n \leq N} \frac{\mu(n)}{n^{s+\frac{1}{2}-a}} \cdot \frac{\log \frac{N}{n}}{\log N} \quad ,$$

where $N = T^{1/2}(\log T)^{-20}$. It is difficult to determine precisely which mollifier will yield the best results; this one is clearly quite good. The last three terms on the right hand side of (1) are easily seen to be $O(\log T)$. Thus it suffices to show that

$$\int_T^{T+U} \log|G\psi(a+it)|dt < (0.33)UR .$$

We are not in a position to deal directly with this integral; instead we use the geometric-arithmetic mean inequality:

$$\int_T^{T+U} \log|G\psi(a+it)|dt \leq \frac{1}{2} U\log\left(\frac{1}{U}\int_T^{T+U} |G\psi(a+it)|^2 dt\right)$$

Here $U = T(\log T)^{-10}$. It now takes Levinson 45 printed pages to show that

$$\int_T^{T+U} |G\psi(a+it)|^2 dt = UF(R) + O(U(\log T)^{-1+}),$$

where

$$(2) \quad F(R) = e^{2R}\left(\frac{1}{2R^3} + \frac{1}{24R}\right) - \frac{1}{2R^3} - \frac{1}{R^2} - \frac{25}{24R} + \frac{7}{12} - \frac{R}{12} .$$

Despite being very long and complicated, this calculation is essentially routine. We take $R = 1.3$, to get $F(R) < 2.3502$, and then we see that

$$\frac{1}{2} \log F(R) < 0.42725 < 0.429 = (0.33) \cdot (1.3) \quad .$$

Note that in our estimate $N_G(T) < \frac{c}{2\pi} T \log T$, if $c > \frac{1}{2}$ we obtain no result. Thus it is necessary to prove (2) before one can see that the method succeeds. This is different from earlier methods, which succeed without the need of calculating constants.

3. SUMS OF INDEPENDENT RANDOM VARIABLES

N.B. I have tried in vain to find Theorems 1 and 2 in the literature. Any useful references would be greatly appreciated.

If $X_1, X_2, \ldots$ are independent random variables and $X = \sum X_k$ converges almost everywhere, we seek to estimate the distribution of X in terms of the distribution functions of the X_k. We restrict our attention to a special case; we let $f(\underline{\theta}) = \sum_{k=1}^{\infty} r_k \sin 2\pi\theta_k$ where $r_k \searrow 0$ and $\sum_{k=1}^{\infty} r_k^2 < \infty$. Here $\underline{\theta} \in \mathbf{T}^\infty$, and our probability measure is Lebesgue measure on $\mathbf{T}^\infty$. The condition that $r_k \searrow 0$ does not occasion any loss of generality, since we may permute k's and translate θ_k's without altering the distribution function of $f(\underline{\theta})$. The condition that $\Sigma r_k^2 < \infty$ is the necessary and sufficient condition that the sum defining $f(\underline{\theta})$ should converge almost everywhere (as we see by Kolmogorov's theorem). (If $\sum r_k^2 = \infty$ then the sum diverges a.e.).

THEOREM 1. Let $f(\underline{\theta})$ be as above. For any integer $K \geq 1$,

$$P(f(\underline{\theta}) \geq 2 \sum_{k=1}^{K} r_k) \leq \exp(-\tfrac{3}{4}(\sum_{k=1}^{K} r_k)^2 (\sum_{k>K} r_k^2)^{-1}),$$

and

$$P(f(\underline{\theta})\geq \frac{1}{2}\sum_{k=1}^{K} r_k) \geq 2^{-40}\exp(-100(\sum_{k=1}^{K} r_k)^2(\sum_{k>K} r_k{}^2)^{-1}) \quad .$$

If the Riemann Hypothesis is true and the imaginary parts $\gamma > 0$ of the zeros are linearly independent then the asymptotic distribution function of $\frac{\psi(e^y)-e^y}{e^{\frac{1}{2}y}}$ is the distribution function of the random variable $g(\theta) = \sum_{\gamma>0} \frac{2}{|\rho|} \sin 2\pi\theta_\rho$. That is, for every real number V,

$$\lim_{Y\to\infty} \frac{1}{Y} \operatorname{meas}\{y\in[0,Y]:\ \psi(e^y)-e^y\geq e^{\frac{1}{2}y} V\} = P(g(\underline{\theta})\geq V) \ .$$

By means of Theorem 1 we obtain rough estimates for this probability: There are constants $0 < c_1 < c_2$ such that for large $V > 0$,

$$\exp(-e^{c_2\sqrt{V}}) \leq P(g(\underline{\theta})\geq V) \leq \exp(-e^{c_1\sqrt{V}}) \quad .$$

We can refine the upper bound by taking more care, but the lower bound presents more difficulties. However, when the r_k decrease very rapidly, as they do here, we can obtain a sharp lower bound from

THEOREM 2. Let $f(\underline{\theta})$ be as in Theorem 1. If δ is so small that

$$\sum_{r_k>\delta} (r_k-\delta) \geq V,$$

then

$$P(f(\underline{\theta})\geq V) \geq \frac{1}{2} \exp\Big(-\frac{1}{2} \sum_{r_k>\delta} \log \frac{\pi^2 r_k}{2\delta}\Big) \quad .$$

Using the above, we can show that there are constants $0 < c_1 < c_2$ such that

$$\exp(-c_2\sqrt{V}\, e^{\sqrt{2\pi V}}) \leq P(g(\theta)>V) \leq \exp(-c_1\sqrt{V}\, e^{\sqrt{2\pi V}}) \quad .$$

This inequality suggests the following

CONJECTURE.

$$\overline{\underline{\lim}} \frac{\psi(x)-x}{x^{\frac{1}{2}}(\log\log\log x)^2} = \pm \frac{1}{2\pi} \quad .$$

To prove Theorem 1 we use the method of the Laplace transform; that is we use the function

$$E(e^{\lambda f}) = \int e^{\lambda f} d\mu \ .$$

Here $e^{\lambda f}$ is a product of factors $e^{\lambda r_k \sin 2\pi\theta_k}$, each depending on only one coordinate, so the integral above is

$$= \prod_{k=1}^{\infty} \int_0^1 e^{\lambda r_k \sin 2\pi\theta} d\theta = \prod_{k=1}^{\infty} I(\lambda r_k)$$

where

(1) $$I(r) = \int_0^1 e^{r\sin 2\pi\theta} d\theta .$$

Before proceeding further, we establish the following basic inequalities:

(2) $$I(r) \leq \begin{cases} e^r \\ e^{\frac{1}{4}r^2} \end{cases} \quad \text{all} \quad r \geq 0 ,$$

and

(3) $$I(r) > \begin{cases} 2e^{\frac{r}{2}} & r \geq 7, \\ e^{\frac{r^2}{19}} & 0 < r \leq 7 \end{cases}$$

The first bound in (2) is trivial, since the integrand in (1) is never more than e^r. (A deeper analysis reveals that $I(r) \sim \frac{e^r}{\sqrt{2\pi r}}$ for $r \to \infty$.) Now $e^{r\sin 2\pi\theta} \geq e^{\frac{\sqrt{3}}{2}r}$ for $\frac{1}{6} \leq \theta \leq \frac{1}{3}$, so that $I(r) \geq \frac{1}{6}e^{\frac{\sqrt{3}}{2}r}$, and this is $>2e^{\frac{1}{2}r}$ for $r \geq 7$. To derive the remaining bounds, which pertain to small values of r, we determine the power series expansion of $I(r)$. Using the power series expansion for e^z, we find that

$$I(r) = \sum_{n=0}^{\infty} \frac{r^n}{n!} \int_0^1 (\sin 2\pi\theta)^n d\theta .$$

Here the integrals vanish for odd n, so that

$$I(r) = \sum_{n=0}^{\infty} \frac{r^{2n}}{(2n)!} \int_0^1 (\sin 2\pi\theta)^{2n} d\theta = \frac{1}{n} \sum_{n=0}^{\infty} \frac{r^{2n}}{(2n)!} \int_0^1 x^{n-\frac{1}{2}} (1-x)^{-\frac{1}{2}} dx \; .$$

$$(x = (\sin 2\pi\theta)^2)$$

This last integral is

$$B(n + \tfrac{1}{2}, \tfrac{1}{2}) = \frac{\Gamma(n+\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(n+1)} = \frac{(2n)!\pi}{2^{2n}(n!)^2} \; ;$$

hence

$$I(r) = \sum_{n=0}^{\infty} \frac{(\frac{r}{2})^{2n}}{(n!)^2} \; .$$

We recognize this as a Bessel function. (I guess we do - the library is closed). As for the remaining inequalities, we see that

$$I(r) < \sum_{n=0}^{\infty} \frac{(\frac{r}{2})^{2n}}{n!} = e^{\frac{r^2}{4}} \; ,$$

and that

$$I(r) \geq 1 + \frac{r^2}{4} \; ;$$

if $0 < r \leq 7$ then this $\geq e^{\frac{r^2}{19}}$.

From the second part of (2) we note that

$$E(e^{\lambda f}) \leq \exp\left(\frac{\lambda^2}{4} \sum_{k=1}^{\infty} r_k^{\,2}\right) < \infty \; ;$$

thus all integrals that we have considered are absolutely convergent and $|f| < \infty$ a.e.; we need Kolmogorov's theorem only to see that the condition $\sum r_k^2 < \infty$ is not unduly restrictive.

With more work we could show that the first part of (3) holds for $r \geq 5$, and that if $0 \leq r \leq 5$ then $I(r) \geq e^{\frac{r^2}{8}}$. This would yield better constants in the lower bound of Theorem 1, but we prefer to give a self-contained account which can be easily verified.

To obtain the upper bound in Theorem 1 we use Chernoff's inequality, which asserts that for any $\lambda > 0$, $V \geq 0$,

$$P(f \geq V) \leq e^{-\lambda V} E(e^{\lambda f}) .$$

To see this we note that

$$P(f \geq V) \leq \int_{f \geq V} d\mu \leq \int e^{\lambda(f-V)} d\mu = e^{-\lambda V} \int e^{\lambda f} d\mu = e^{-\lambda V} E(e^{\lambda f}) .$$

For the lower bound, we use the inequality

$$(4) \qquad P(F \geq \tfrac{1}{2}E(F)) \geq \tfrac{1}{4}E(F)^2/E(F^2),$$

valid for any non-negative F. To see this, we first note that

$$\int_{F \leq \frac{1}{2}E(F)} F d\mu \leq \frac{1}{2}E(F) ,$$

so that

$$\int_{F \geq \frac{1}{2}E(F)} F d\mu \leq \frac{1}{2}E(F) .$$

Hence

$$E(F)^2 = (\int F \, d\mu)^2 \leq (2 \int_{F \geq \frac{1}{2}E(F)} F \, d\mu)^2 ,$$

and by Cauchy's inequality this is

$$\leq 4 \int_{F \geq \frac{1}{2}E(F)} 1 \quad d\mu \int_{F \geq \frac{1}{2}E(F)} F^2 d\mu \leq 4P(F \geq \frac{1}{2}E(F))E(F^2) .$$

We shall apply this with $F = e^{\lambda f}$.

We now proof the upper bound in Theorem 1. By (2) we have

$$E(e^{\lambda f}) = \prod I(\lambda r_k) \leq \exp(\lambda \sum_{r_k \leq K} r_k + \frac{\lambda^2}{4} \sum_{k>K} r_k{}^2) .$$

We now take

$$\lambda = (\sum_{k=1}^{K} r_k) / \sum_{k>K} r_k{}^2 , \quad V = 2 \sum_{k=1}^{K} r_k .$$

Then by Chernoff's inequality

$$P(f \geq 2\sum_{k=1}^{K} r_k) \leq e^{-\lambda V} E(e^{\lambda f}) = \exp(- \frac{3}{4}(\sum_{k=1}^{K} r_k)^2 / \sum_{k>K} r_k{}^2);$$

this is the desired result.

For the lower bound we use (4) with $F = e^{\lambda f}$, and we take $\lambda > 0$ so that

(5) $$E(e^{\lambda f}) = 2\exp(\frac{\lambda}{2}\sum_{k=1}^{K} r_k) .$$

Thus $F \geq \frac{1}{2}E(F)$ if and only if (and only if and only if) $f \geq \frac{1}{2}\sum_{k=1}^{K} r_k$. Before proceeding further we demonstrate that there is a λ such that (5) holds. The two sides of (5) are continuous functions of λ; for $\lambda = 0$ the left hand side is $=1$ while the right hand side is $=2$. Moreover, by (3) we see that if $\lambda > \frac{7}{r_k}$ then

$$E(F) \geq \prod_{k=1}^{K} I(\lambda r_k) \geq 2^K \exp(\frac{1}{2}\lambda\sum_{k=1}^{K} r_k) \geq 2\ \exp(\frac{\lambda}{2}\sum_{k=1}^{K} r_k) .$$

Thus there is such a λ, and

(6) $$\lambda \leq 7/r_k$$

Hence

$$E(F) > \prod_{k>K} I(\lambda r_k) > \exp(\frac{\lambda^2}{19}\sum_{k>K} r_k^{\ 2}) .$$

This with (5) gives a quadratic inequality for λ,

(7) $$\frac{\lambda^2}{19}\sum_{k>K} r_k^{\ 2} \leq \frac{\lambda}{2}\sum_{k\leq K} r_k + \log 2 .$$

But if $a\lambda^2 \leq b\lambda + c$ with $a,b,c \geq 0$ then

$$\lambda \leq \frac{b+\sqrt{b^2+4ac}}{2a} = \frac{b}{2a}\left(1+\sqrt{1+\frac{4ac}{b^2}}\right) < \frac{b}{2a}\left(1+1+\frac{2ac}{b^2}\right) .$$

Thus

$$\text{(8)} \qquad \lambda \leq \frac{19}{2} \frac{\sum_{k=1}^{K} r_k}{\sum_{k>K} r_k^2} + \frac{2 \log 2}{\sum_{i=1}^{K} r_k}$$

Finally, by (5) we have

$$\tfrac{1}{4}E(F)^2/E(F^2) = \exp\left(\lambda\sum_{k=1}^{K} r_k\right)/E(F^2) ,$$

and by (2) this is

$$\geq \exp\left(-\lambda \sum_{k=1}^{K} r_k - \lambda^2 \sum_{k>K} r_k^2\right) .$$

By (7) this is

$$\geq 2^{-19}\exp\left(-\frac{21}{2}\lambda \sum_{k=1}^{K} r_k\right),$$

and by (8) this is

$$\geq 2^{-40}\exp\left(-\frac{21.19}{2}\left(\sum_{k=1}^{K} r_k\right)^2 / \sum_{k>K} r_k^2\right),$$

which gives the desired bound.

The above proof was suggested to me by Andrew

Odlyzko. My original (much more complicated) proof depended on moment inequalities,

$$(9) \qquad \int \Big|\sum_{k=1}^{\infty} r_k e(\theta_k)\Big|^{2n} d\mu \le n!\Big(\sum_{k=1}^{\infty} r_k^{\,2}\Big)^n .$$

A corresponding lower bound

$$(10) \qquad \int \Big|\sum_{k=1}^{\infty} r_k e(\theta_k)\Big|^{2n} d\mu \ge 2^{-n} n!\Big(\sum_{k=1}^{\infty} r_k^{\,2}\Big)^2$$

holds provided that

$$(11) \qquad \sum_{k<n} r_k^{\,2} \le \tfrac{1}{2} \sum_{k=1}^{\infty} r_k^{\,2} .$$

It is not difficult to derive an upper bound for $P(|f|\ge V)$ from (9), but the condition (11) restricts our use of (10). We use (10) if $\left(\sum_{k=1}^{K} r_k\right)^2 \le \frac{1}{3}K \sum_{k>K} r_k^{\,2}$. Otherwise we use a less refined form of Theorem 2, which asserts that

$$P\Big(\Big|\sum_{k=1}^{\infty} r_k e(\theta_k)\Big| \ge \tfrac{1}{2}\sum_{k=1}^{K} r_k\Big) \ge 3^{-K} .$$

One advantage of the moment method is that we can deal directly with a complex-valued random variable.

We now prove Theorem 2. We break the sum defining f into two ranges, $f = \sum_{r_k>\delta} + \sum_{r_k\le\delta} = f_1 + f_2$. If $f_1 \ge V$ and $f_2 \ge 0$ then $f \ge V$. Hence

$P(f \geq V) \geq P(f_1 \geq V) P(f_2 \geq 0) = \frac{1}{2} P(f_1 \geq V)$. Since $\cos x \geq 1 - \frac{x^2}{2}$ for all x, if $|\theta_k - \frac{1}{4}| \leq (\frac{\delta}{2\pi^2 r_k})^{1/2}$ then

$$\sin 2\pi\ \theta k \geq \cos(\frac{2\delta}{r_k})^{1/2} \geq 1 - \frac{\delta}{r_k} .$$

Let $B \subseteq \mathbf{T}^\infty$ be the box for which $|\theta_k - \frac{1}{4}| \leq (\frac{\delta}{2\pi^2 r_k})^{1/2}$ for all k such that $r_k > \delta$. For $\underline{\theta} \in B$ we have

$$f_1(\underline{\theta}) = \sum_{r_k > \delta} r_k \sin 2\pi\ \theta_k \geq \sum_{r_k > \delta} r_k (1 - \frac{\delta}{r_k}) ,$$

and by hypothesis this last quantity is $\geq V$. Thus $P(f_1 \geq V) \geq$ meas B, and it suffices to note that

$$B = \prod_{r_k > \delta} (\frac{2\delta}{\pi^2 r_k})^{\frac{1}{2}} = \exp(-\frac{1}{2} \sum_{r_k > \delta} \log \frac{\pi^2 r_k}{2\delta}) .$$

4. VAUGHAN'S VERSION OF VINOGRADOV'S METHOD

We develop a method for estimating sums of the sort $\sum_{p\le X} f(p)$, or equivalently, $\sum_{n\le X} \Lambda(n)f(n)$. The method fails if f is multiplicative, but it can be used to estimate averages $\sum_{j=1}^{J} \left|\sum_{p\le X} f_j(p)\right|$ over multiplicative functions. In this way Vaughan has given a simple proof of Bombieri's theorem. This is also the method used by Heath-Brown and Patterson to settle Kummer's conjecture on the cubic Gaussian sum.

Let $P = \prod_{p\le\sqrt{X}} p$. The sieve of Eratosthenes asserts that the number n, $1 < n \le X$ is prime if and only if $(n,P) = 1$. Hence Vinogradov started by writing

$$f(1) + \sum_{X^{\frac{1}{2}}<p\le X} f(p) = \sum_{\substack{n\le X \\ (n,P)=1}} f(n) = \sum_{n\le X} f(n) \sum_{\substack{t\mid n \\ t\mid P}} \mu(t) =$$

$$= \sum_{\substack{t\mid P \\ t\le X}} \mu(t) \sum_{r\le \frac{X}{t}} f(rt)$$

Thus we are led to estimate sums of the sort $\sum_{r\le X/t} f(rt)$. We expect such a sum to be $o(\frac{X}{t})$ when $\frac{X}{t}$ is large. However, when t is nearly as large as X we must perform further regrouping of terms to obtain cancellation.

We now present Vaughan's decomposition of $\Lambda(n)$.

Let

$$F(s) = \sum_{m \leq U} \frac{\Lambda(m)}{m^s}, \quad G(s) = \sum_{d \leq V} \frac{\mu(d)}{d^s}.$$

Then

$$-\frac{\zeta'}{\zeta}(s) = F(s) - \zeta(s)F(s)G(s) - \zeta'(s)G(s) + \left(-\frac{\zeta'}{\zeta}(s) - F(s)\right)(1 - \zeta(s)G(s))$$

for $\sigma > 1$. We compare the Dirichlet series coefficients of these functions, and find that

$$\Lambda(n) = a_1(n) + a_2(n) + a_3(n) + a_4(n),$$

where

$$a_1(n) = \begin{cases} \Lambda(n) & n \leq U, \\ 0 & n > U; \end{cases}$$

$$a_2(n) = -\sum_{\substack{mdr=n \\ m \leq U \\ d \leq V}} \Lambda(m)\mu(d);$$

$$a_3(n) = \sum_{\substack{hd=n \\ d \leq V}} \mu(d) \log h;$$

and

$$a_4 = -\sum_{\substack{mk=n \\ m>U \\ k>1}} \Lambda(m)\Big(\sum_{\substack{d|k \\ d\leq V}} \mu(d)\Big).$$

We multiply by $f(n)$ and sum to find that

$$S = \sum_{n\leq X} \Lambda(n)f(n) = S_1 + S_2 + S_3 + S_4,$$

where $S_i = \sum_{n\leq X} a_i(n)f(n)$.

Although it is not necessary for the method we assume henceforth that

$$|f(n)| = 1$$

for all n; this helps us to gauge the power of the method. The trivial bound for S is thus $\ll X$; we seek a bound which is $o(X)$.

The sum S_1 we bound trivially:

(1) $$S_1 \ll U .$$

The sum S_2 we write as

$$S_2 = -\sum_{t\le UV}\Big(\sum_{\substack{t=md\\ m<U\\ d\le V}} \mu(t)\Lambda(m)\Big)\sum_{r\le \frac{X}{t}} f(rt).$$

Note that we again have a linear combination of the sums $\sum_{r\le X/t} f(rt)$, but that now we can control the range of t by ensuring that UV is rather smaller than X. As

$$\sum_{m|t} \Lambda(m) = \log t \le \log UV,$$

we see that

$$S_2 \ll (\log UV)\sum_{t\le UV} \Big|\sum_{r\le X/t} f(rt)\Big| . \tag{2}$$

The sum S_3 is of the same form, since

$$S_3 = \sum_{d\le V}\mu(d)\sum_{h\le X/d} f(dh)\log h = \sum_{d\le V}\mu(d)\sum_{h\le X/d} f(dh)\int_1^h \frac{dw}{w}$$

$$= \int_1^X \sum_{d\le V}\mu(d)\sum_{w\le h\le X/d} f(dh)\frac{dw}{w} .$$

Hence

$$S_3 \ll (\log X)\sum_{d\le V} \max\Big|\sum_{w\le h\le X/d} f(dh)\Big| . \tag{3}$$

The sum S_4 is more complicated. We note that $\sum_{\substack{d|k \\ d<V}} \mu(d) = 0$ for $1 < k \le V$, so that

$$S_4 = \sum_{U<m\le\frac{X}{V}} \Lambda(m) \sum_{V<k\le\frac{X}{m}} \Big(\sum_{\substack{d|k \\ d\le V}} \mu(d)\Big) f(mk).$$

Suppose that $\Delta = \Delta(f,M,X,V)$ is such that

$$\Big|\sum_{M<m\le 2M} b_m \sum_{V<k\le\frac{X}{m}} c_k f(mk)\Big| \le \Delta\Big(\sum_{M}^{2M} |b_m|^2\Big)^{\frac{1}{2}}\Big(\sum_{V<k\le\frac{X}{M}} |c_k|^2\Big)^{\frac{1}{2}}$$

for all complex numbers b_m, c_k. Such bilinear forms are familiar, and we are equipped to estimate Δ. Then we see that

$$
\begin{aligned}
S_4 &\ll (\log X) \max_{U\le M\le\frac{X}{V}} \Delta\Big(\sum_{M}^{2M} \Lambda(m)^2\Big)^{\frac{1}{2}}\Big(\sum_{k\le\frac{X}{M}} d(k)^2\Big)^{\frac{1}{2}} \\
&\ll X^{\frac{1}{2}}(\log X)^3 \max_{U\le M\le\frac{X}{V}} \Delta\ . \qquad (4)
\end{aligned}
$$

We now pause to examine how much we have lost. If $f \equiv 1$ we obtain the trivial bounds, $S_1 \ll U$, $S_2 \ll X(\log UV)^2$, $S_3 \ll X(\log XV)^2$. In this case we also have the trivial bound $\Delta \ll X^{\frac{1}{2}}$, which gives $S_4 \ll X(\log X)^3$. Hence if we can obtain some cancellation in the sums

$\sum_{r \le X/t} f(rt)$ then $S_2, S_3 = o(X)$, while a good estimate of Δ gives $S_4 = o(X)$. Note, however, that if f is totally multiplicative then from the choice $b_m = \overline{f(m)}$, $c_k = \overline{f(k)}$ we see that $\Delta \approx X^{1/2}$, and the method fails. To estimate Δ we use Cauchy's inequality to see that the bilinear form above is

$$\le \Big(\sum_{M}^{2M} |b_m|^2\Big)^{\frac{1}{2}} \Big(\sum_{M}^{2M} \Big| \sum_{V<k\le X/M} c_k f(mk)\Big|^2\Big)^{\frac{1}{2}}.$$

Here the second sum over m is

$$= \sum_{V<k\le \frac{X}{M}} c_k \sum_{V<\ell\le \frac{X}{M}} \overline{c_\ell} \sum_{\substack{M<m<2M \\ m\le X/k \\ m\le X/\ell}} f(mk)\overline{f(m\ell)}.$$

We note that $|c_k \overline{c_\ell}| \le \frac{1}{2}|c_k|^2 + \frac{1}{2}|c_\ell|^2$; hence the above is

$$\ll \sum_{V<k\le \frac{X}{M}} |c_k|^2 \sum_{V<\ell\le \frac{X}{M}} \Big| \sum_{\substack{M<m<2M \\ m\le X/k \\ m\le X/\ell}} f(mk)\overline{f(m\ell)}\Big| .$$

Thus we find that

$$(5) \qquad \Delta \ll \Big(\max_{V<k\le \frac{X}{M}} \sum_{V<\ell\le \frac{X}{M}} \Big| \sum_{\substack{M<m<2M \\ m\le X/k \\ m\le X/\ell}} f(mk)\overline{f(m\ell)}\Big|\Big)^{\frac{1}{2}} .$$

If $f \equiv 1$ we again obtain the trivial estimate $\Delta \ll X^{\frac{1}{2}}$.

Combining estimates (1) - (5), we see that we have proved that if $U \geq 2$, $V \geq 2$, $UV \leq X$, $|f(n)| \leq 1$, then

$$\sum_{n \leq X} f(n)\Lambda(n) \ll U + (\log X) \sum_{t \leq UV} \max_{w} \Big| \sum_{w \leq r \leq X/t} f(rt) \Big|$$

$$+ X^{\frac{1}{2}} (\log X)^3 \max_{U \leq M \leq \frac{X}{V}} \max_{V < k \leq \frac{X}{M}} \Big(\sum_{V < \ell \leq \frac{X}{M}} \Big| \sum_{\substack{M \leq m < 2M \\ m \leq X/k \\ m \leq X/\ell}} f(mk)\overline{f(m\ell)} \Big| \Big)$$