On the Distribution of Prime Numbers:

Michael Emanuel

May 5, 1998

Indroduction:

We are all familiar with Dirichlet's celebrated theorem that if $(a, q) = 1$, then there are inifinitely many primes congruent to a mod q. Furthermore, by expanding the methods used by Hadamard and de la Vallée Poussin to prove the Prime Number Theorem, one can easily show that each residue class has approximately "the same" number of primes. That is, one can show that the prime counting functions $\pi(x, q, a)$ behave like:

$$\pi(x, q, a) \sim \frac{\pi(x)}{\varphi(q)}$$

But it has long been empirically known that more primes are quadratic nonresidues than are quadratic residues mod q for small q. Chebyshev first noted this trend in 1853, so we refer to it as "Chebyshev's Bias" in his honor. In this paper, we will investigate this phenomenon of biased distributions of primes to a given modulus. In order to make any real progess, however, we must assume two as-yet unproven conjectures, the Generalized Riemann Hypothesis and the Grand Simplicity Hypothesis:

Generalized Riemann Hypothesis (GRH):
The non-trivial zeroes of any Dirichlet L-function $L(s, \chi)$ all have real part one half.

Grand Simplicity Hypothesis (GSH):
The set of $\gamma \geq 0$ such that $L(\frac{1}{2} + i\gamma) = 0$ is linearly independent over $\mathbb{Q}$.

Under these very natural assumptions, we will first prove the existence of the "Chebyshev Bias" toward quadratic nonresidues using only the Riemann Hypothesis, and then introduce a limiting distribution μ that measures the magnitude and direction of the bias. In the next section, we will use the Grand Simplicity Hypothesis to develop a very pretty product formula for its Fourier transform $\hat{\mu}$. Although the computation is somewhat gory, the resulting formula is worth the trouble. The third and final section explores some of the far-reaching implications of this product formula. In particular, we work out four interesting applications. First, we will prove that there is always a measurable bias towards quadratic nonresidues. Second, we will show that as q tends to infinity, the bias in primes mod q dissolves. Third, while the bias goes away as q becomes large, we will see that it is only zero in two simple cases. Finally, we will demonstrate an efficient numerical procedure for estimating the logarithmic densities of the sets on which the quadratic nonresidues are beating out the quadratic residues for sequences mod3, 4, 5, 7, 11, and 13. A simple adaptation of this technique allows us to estimate the density of the set on which $\pi(x) > \mathrm{Li}(x)$.

Existence of the Chebyshev Bias:

We recall the definition of the logarithmic density of a set P of integers:

$$\delta(P) = \lim_{M\to\infty} \frac{1}{\log M} \int_{P\cap[2,M]} \frac{dt}{t}$$

provided that this limit exists.
Now the fact that the logarithmic densities of the sets

$$\{x \in \mathbb{Z} : x \text{ is prime and } x \equiv a \pmod{q}\}$$

are equal when $(a, q) = 1$ in no way means that there isn't a meaningful bias; the logarithmic density is simply too crude a tool to measure this behavior. To get around this problem, we use the finer measurement of the so-called "primes race": given $a_1, \ldots, a_r \in (\mathbb{Z}/q\mathbb{Z})^*$, we consider the set

$$P_{q;a_1,\ldots,a_r} := \{x \in \mathbb{R} \mid \pi(x, q, a_1) > \pi(x, q, a_2) > \cdots > \pi(x, q, a_r)$$

So $P_{q;a_1,\ldots,a_r}$ is the set of numbers where the "scorecard of the primes race" is $a_1, \ldots, a_r$. Our goal is to develop the tools to estimate the logarithmic densities of these sets. To that end, we define the vector-valued function

$$E_{q;a_1,\ldots,a_r}(x) = \frac{\log x}{x} \times (\varphi(q)\pi(x, q, a_1) - \pi(x), \ldots, \varphi(q)\pi(x, q, a_r) - \pi(x))$$

We can see that the term $\varphi(q)\pi(x, q, a_j) - \pi(x)$ measures the "surplus" of primes $p \equiv a \pmod{q}$, since by Dirichlet's theorem this quantity should be near zero; the ${}^{\log x}/_{x}$ term is a normalization factor. Our first task is to explain the origin of the bias towards quadratic nonresidues. Recall the definition of the ψ function:

$$\psi(x, \chi) := \sum_{n\leq x} \chi(n)\Lambda(n)$$

where χ is a Dirichlet character and Λ is the von Mangoldt function $\Lambda(n) = \log p$ if n is a prime power $n = p^k$, and zero otherwise. Now the contour integral formula for $\psi(x,\chi)$ tells us that if $\chi \neq \chi_0$, $x \geq 2$, and $T \geq 1$, then

$$\psi(x,\chi) = \sum_{|\gamma_\chi| \leq T} \frac{x^\rho}{\rho} + O\left(\frac{x\log^2(xT)}{T} + \log x\right)$$

Here ρ runs over the non-trivial zeroes of $L(x,\chi)$ (i.e. the zeroes in the critical strip $0 \leq \mathrm{Re}(s) \leq 1$) But we are assuming the Riemann Hypothesis, so the real part of all these zeroes is one half and

$$\psi(x,\chi) = \sqrt{x} \sum_{|\gamma_\chi| \leq T} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + O\left(\frac{x\log^2(xT)}{T} + \log x\right)$$

Now we recall that the characteristic function $\mathbf{1}_a$ for $a \bmod q$ is given by $\mathbf{1}_a(n) = \frac{1}{\varphi(q)} \sum_\chi \bar{\chi}(a)\chi(n)$, so

$$\psi(x,q,a) = \sum_{\substack{n \leq x \\ n \equiv a(q)}} \Lambda(n) = \frac{1}{\varphi(q)} \sum_{\chi \bmod q} \bar{\chi}(a) \sum_{n \leq x} \Lambda(n)\chi(n) = \frac{1}{\varphi(q)} \sum_{\chi \bmod q} \bar{\chi}(a)\psi(x,\chi)$$

When we proved the Prime Number Theorem, we did so by first showing that $\psi(x)$ behaved like x and then made the observation that in the sum

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{p^k \leq n} \log p$$

the sum is dominated by the term involving only p^1; that is, we showed that

$$\sum_{p \leq n} \log p \ggg \sum_{\substack{p^k \geq n \\ k \geq 2}} \log p$$

The idea in proving the Chebyshev bias is similar, but this time we will need to consider *both* the linear and square terms in the sum for $\psi(x,\chi)$. It is the square term that introduces the discrepancy between quadratic residues and nonresidues. So we define

$$E(x,q,a) = (\varphi(q)\pi(x,q,a) - \pi(x))\frac{\log x}{\sqrt{x}}$$

Note that $E(x,q,a)$ is just a convenient notation for $E_{q;a}(x)$, so it measures the excess of primes $p \equiv a \pmod q$ over the Dirichlet quota $\pi(x)/\varphi(q)$. We are now ready to prove the main lemma.

Lemma: *As $x \to \infty$ we have*

$$E(x,q,a) = -c(q,a) + \sum_{\chi \neq \chi_0} \bar{\chi}(a) \frac{\psi(x,\chi)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right)$$

where the term $c(q,a)$ is defined to be

$$c(q,a) = -1 + \sum_{\substack{b^2 \equiv a(q) \\ 0 \leq b \leq q-1}} 1$$

It is the constant term $-c(q,a)$ that accounts for the bias towards quadratic nonresidues.

Proof:

Let $\theta(x,q,a) = \sum_{\substack{p \leq x \\ p \equiv a(q)}} \log p$. Then

$$\pi(x,q,a) = \int_2^x \frac{d\theta(t,q,a)}{\log t}$$

This looks a bit obscure at first, but it makes sense since $d\theta$ is effectively the sum of Dirac delta functions of weight $\log p$ at each prime $p \leq x$ that is also congruent to a, so each of them contributes 1 in the integral. Now by two applications of Dirichlet's Theorem in its incarnation

$$\psi(x,q,a) = \frac{x}{\varphi(q)} + O\left(\sqrt{x} \log^2 x\right)$$

(the error is smaller than usual here because we are assuming the Riemann Hypothesis!) we compute that

$$\psi(x,q,a) = \theta(x,q,a) + \Big(\sum_{b^2 \equiv a(q)} 1\Big) \frac{\sqrt{x}}{\varphi(q)} + O\left(\frac{\sqrt{x}}{\log x}\right)$$

Here is the computation:

$$\psi(x,q,a) = \sum_{\substack{n \leq x \\ n \equiv a}} \Lambda(n) = \sum_{\substack{p^k \leq x \\ p^k \equiv a}} \log p = \sum_{\substack{p \leq x \\ p \equiv a}} \log p + \sum_{\substack{p^2 \leq x \\ p^2 \equiv a}} \log p + \sum_{\substack{p^k \leq x; k \geq 3 \\ p^k \equiv a}} \log p$$

The first term (i.e. the sum over p) is just $\theta(x,q,a)$ by definition. The last term is an error which can be absorbed into $O(\sqrt{x}/\log x)$ since

$$\sum_{\substack{p^k \leq x; k \geq 3 \\ p^k \equiv a}} \log p \leq \sum_{p \leq \sqrt[3]{x}} \log p = O\left(x^{1/3} \log x\right)$$

The second term is:

$$\sum_{\substack{p^2\le x\\ p^2\equiv a}} \log p = \sum_{\substack{p\le\sqrt{x}\\ p^2\equiv a}} \log p = \sum_{b^2\equiv a}\Big(\sum_{\substack{p\le\sqrt{x}\\ p\equiv b}} \log p\Big)$$

Now the inner sum $\sum\limits_{\substack{p\le\sqrt{x}\\ p\equiv b}} \log p$ differs from $\psi(\sqrt{x}, q, a)$ by a factor

$$\sum_{\substack{p^k\le\sqrt{x};k\ge2\\ p\equiv b}} \log p \ll x^{1/4}\log x$$

So we can replace it with $\sqrt{x}/\varphi(q)$ without compromising our error estimate of $O\left(\sqrt{x}/\log x\right)$. Thus we consolidate the second term to be

$$\sum_{b^2\equiv a}\Big(\sum_{\substack{p^2\le x\\ p^2\equiv a}} \log p\Big) = \sum_{b^2\equiv a}\frac{\sqrt{x}}{\varphi(q)}$$

and we get the desired result.
Solving for $\theta(x,q,a)$ we find that

$$\theta(x,q,a) = \psi(x,q,a) - \Big(\sum_{b^2\equiv a(q)} 1\Big)\frac{\sqrt{x}}{\varphi(q)} + O\left(\frac{\sqrt{x}}{\log x}\right)$$

But $\psi(x,q,a) = \frac{1}{\varphi(q)}\sum\limits_{\chi \bmod q} \bar{\chi}(a)\psi(x,\chi)$; so plowing forward with this integral,

$$\pi(x,q,a) = \frac{1}{\varphi(q)}\int_2^x \frac{d\psi(t)}{\log t} + \frac{1}{\varphi(q)}\sum_{\chi\ne\chi_0}\bar{\chi}(a)\int_2^x\frac{d\psi(t,\chi)}{\log t} - \frac{1}{\varphi(q)}\sum_{b^2\equiv a(q)} 1\frac{\sqrt{x}}{\log x}$$

again with error $O(\sqrt{x}/\log x)$. The terms with $\int_2^x \frac{d\psi(t)}{\log t}$ can be easily handled:

$$\psi(x) = \sum_{p^k\le x}\log p = \sum_{p\le x}\log p + \sum_{p^2\le x}\log p + \sum_{p^k\le x;k\ge3}\log p$$

whence $d\psi(x) = \log x\, d\pi(x) + \log(\sqrt{x})d\pi(\sqrt{x}) + O\left(\log x\, d\pi(x^{1/3})\right)$

$$\begin{aligned}\int_2^x \frac{d\psi(t)}{\log t} &= \int_2^x \frac{\log t\, d\pi(t)}{\log t} + \int_2^x \frac{\log(t^{1/2})d\pi(t^{1/2})}{\log t} + O(x^{1/3}) \\ &= \int_2^x d\pi(t) + \int_2^x \frac{1}{2} d\pi(t^{1/2}) = \pi(x) + \frac{\sqrt{x}}{\log x} + O(x^{1/3})\end{aligned}$$

Now we use this to simplify our expression for $\pi(x, q, a)$:

$$\begin{aligned}\pi(x,q,a) = \frac{1}{\varphi(q)}\left(\pi(x) + \frac{\sqrt{x}}{\log x}\right) + \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} \bar{\chi}(a) \frac{\psi(x,\chi)}{\log x} \\ -\frac{1}{\varphi(q)}\Big(\sum_{b^2 \equiv a(q)} 1\Big)\frac{\sqrt{x}}{\log x} + O\left(\sum_{\chi \neq \chi_0} \left|\int_2^x \frac{\psi(t,\chi)\,dt}{t\log^2 t}\right| + \frac{\sqrt{x}}{\log^2 x}\right)\end{aligned}$$

Consolidating the $\sqrt{x}/\log x$ terms once more this becomes

$$\pi(x,q,a) = \frac{\pi(x)}{\varphi(q)} + \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} \bar{\chi}(a) \frac{\psi(c,\chi)}{\log x} - \frac{c(q,a)}{\varphi(q)} \frac{\sqrt{x}}{\log x}$$

All that now remains is to show that our error term is in fact $O(\sqrt{x}/\log x)$. Let

$$G(x,\chi) = \int_2^x \psi(t,\chi)\,dt$$

By the contour-integral formula for $\psi(x,\chi)$,

$$\psi(x,\chi) = -\sqrt{x} \sum_{|\gamma| \leq T} \frac{x^{i\gamma}}{\frac{1}{2} + i\gamma} + O\left(\frac{x \log^2 T}{T} + \log x\right)$$

We would like to integrate this sum term by term and allow $T \to \infty$ to conclude that:

$$G(x,\chi) = -\sum_{|\gamma| \leq T} \frac{x^{3/2+i\gamma}}{(\frac{1}{2}+i\gamma)(\frac{3}{2}+i\gamma)} + O(x \log x)$$

This is a valid operation, because the sum over γ converges absolutely. To see this, we use the formula for the aymptotic number of zeroes:

$$N(T) = \#\{|\gamma| \leq T\} = \frac{T}{\pi} \log\left(\frac{qT}{2\pi}\right) - \frac{T}{\pi} + O(\log T + \log q)$$

So the summand has absolute value $\sim \, x^{3/2} \, / 1 + \gamma^2$ and occurs with density $dN(t) \sim \log t$. So we show absolute convergence with the related integral

$$\int_{t=1}^{\infty} \frac{x^{3/2} \log t \, dt}{1+t^2} = O(x^{3/2})$$

Now we return to our original error estimate, integrate it by parts, and use this bound on $G(x, \chi)$:

$$\int_2^x \frac{\psi(t,\chi)\,dt}{t \log^2 t} = \frac{G(x,\chi)}{x \log^2 x} - \int_2^x G(t,\chi)\, d(t \log^2 t)^{-1} \ll \frac{x^{3/2}}{x \log^2 x} = \frac{\sqrt{x}}{\log^2 x}$$

So consolidating all our errors, we find that the error is indeed $O(\sqrt{x}/\log^2 x)\Box$

It is more useful to express $E(x, q, a)$ as a finite sum $E^T(x, q, a)$ over zeroes $\frac{1}{2} + i\gamma$ such that $|\gamma| \leq T$ with an error depending on T; so for $T \geq 1$ and $2 \leq x \leq X$:

$$E(x,q,a) = -c(q,a) - \sum_{\chi \neq \chi_0} \bar{\chi}(a) \sum_{|\gamma_\chi| \leq T} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + \varepsilon_a(x,T,X)$$

where

$$\varepsilon_a(x,T,X) = -\sum_{\chi \neq \chi_0} \bar{\chi}(a) \sum_{T \leq |\gamma_\chi| \leq X} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + O_q\left(\frac{\sqrt{x}\log^2 X}{X} + \frac{1}{\log x}\right)$$

Now there are some technical details to be ironed out at this point, but space does not permit us to handle them here. The long and short of it is that by using the approximations E^T and by using some results involving quasiperiodic functions, we can prove that $E_{q;a_1,\ldots,a_r}$ has a limiting distribution $\mu_{q;a_1,\ldots,a_r}$ in the following

Theorem: *If we assume the Riemann Hypothesis, then* $E_{q;a_1,\ldots,a_r}$ *has a limiting distribution* $\mu_{q;a_1,\ldots,a_r}$ *such that for any bounded, continuous function* $f : \mathbb{R}^r \to \mathbb{R}$,

$$\lim_{X\to\infty} \frac{1}{\log X} \int_2^x f\left(E_{q;a_1,\ldots,a_r}(x)\right) \frac{dx}{x} = \int_{\mathbb{R}^r} f(x)\, d\mu_{q;a_1,\ldots,a_r}(x)$$

The crux of the proof is the fact that if A is the closure in the N-torus T^N of the one-parameter subgroup $\Gamma(y) = \{(\gamma_1 y/2\pi, \ldots, \gamma_N/2\pi)\}$, then by the Kronecker-Weyl Theorem $\Gamma(y)$ is equidistributed in A.

We note that once we know what $\mu_{q;a_1,\dots,a_r}$ is, we know essentially everything about the Chebyshev Bias. For example, we can compute the logarithmic density of $P_{q;a_1,\dots,a_r}$:

$$\delta(P_{q;a_1,\dots,a_r}) = \lim_{x\to\infty}\frac{1}{\log x}\int_2^x \mathbf{1}_{P_{q;a_1,\dots,a_r}}\frac{dt}{t} = \int_{\mathbb{R}^r}\mathbf{1}_{P_{q;a_1,\dots,a_r}}\,d\mu$$

where $\mathbf{1}_{P_{q;a_1,\dots,a_r}}$ is the indicator function for $P_{q;a_1,\dots,a_r}$. Integrating over this is the same as integrating over the region where $x_1 > x_2 > \cdots > x_r$, so

$$\delta(P_{q;a_1,\dots,a_r}) = \int_{\{x\in\mathbb{R}^r\,:\,x_1>x_2>\cdots>x_r\}} d\mu = \mu(\{x\in\mathbb{R}^r : x_1 > x_2 > \cdots > x_r\})$$

The Product Formula for $\hat{\mu}$:

Now we develop a surprisingly simple formula for $\hat{\mu}$ under the Grand Simplicity Hypothesis. We recall that under the GSH, the set of $\gamma : L(\frac{1}{2}+i\gamma,\chi) = 0$ is linearly independent over $\mathbb{Q}$. We have already seen that E^T is a good approximation to E, where

$$E^T(x,q,a) = -c(q,a) - \sum_{\chi\neq\chi_0}\bar{\chi}(a)\sum_{|\gamma_\chi|\le T}\frac{x^{i\gamma_\chi}}{\frac{1}{2}+i\gamma_\chi}$$

In fact, E^T tends to E in the limit as $T\to\infty$. Let $\gamma_1,\dots,\gamma_N$ be the imaginary parts of zeroes $\frac{1}{2}+i\gamma$ where $0<\gamma_1<\gamma_2<\cdots<\gamma_N<T$. By combining the zeroes $\frac{1}{2}\pm i\gamma$ and making the substitution $y=\log x$, we simplify the notation by writing

$$E^T(y) = 2\,\mathrm{Re}\Big(\sum_{j=1}^N b_j e^{i\gamma_j y}\Big) + b_0$$

where $b_0 = -(c(q,a_1),\dots,c(q,a_r))$ and $b_j = \left(\dfrac{\bar{\chi}_j(a_1)}{\frac{1}{2}+i\gamma_j},\dots,\dfrac{\bar{\chi}_j(a_r)}{\frac{1}{2}+i\gamma_j}\right)$

So our limiting distribution μ is equal to

$$\mu(f) = \lim_{y\to\infty}\frac{1}{y}\int_{\log 2}^{y} f\Big(2\,\mathrm{Re}\sum_{j=1}^N b_j e^{i\gamma_j y} + b_0\Big)\,dy$$

in the limit as $N\to\infty$. This is the critical stage where we make strong use of GSH. This integral would normally be impossible to evaluate, but since the γ's are linearly independent over $\mathbb{Q}$, the set of $y(\gamma_1,\dots,\gamma_N)$ is equidistributed in the N-torus. This enables us to apply the Weyl theorem in its integral form and conclude that

$$\lim_{M\to\infty} \frac{1}{M} \int_{\log 2}^{M} f\Big(\sum_{j=1}^{N} \mu_j(y) + b_0\Big)\, dy =$$

$$\lim_{\substack{M_i\to\infty \\ 0\le i\le N}} \frac{1}{\prod M_i} \int_{\log 2}^{M_1} \int_{\log 2}^{M_2} \cdots \int_{\log 2}^{M_N} f\Big(\sum_{j=1}^{N} \mu_j(y_j) + b_0\Big)\, \mathbf{dy}$$

where we have made the substitution $\mu_j(y) = 2\,\mathrm{Re}\left(\sum_{k=1}^{N} b_j e^{i\gamma_k y}\right)$. Before we can compute $\hat{\mu}$, we recall how the Fourier transform is defined on an arbitrary distribution: let $\langle D, \varphi\rangle$ represent the action of the distribution D on the test-function φ. Then the Fourier transform $\mathcal{F}$ of a distribution D is defined by

$$\langle \mathcal{F}D, \varphi\rangle = \langle D, \mathcal{F}\varphi\rangle$$

by the usual method in distribution theory of "passing the operator to the test function." But we want to think of $\hat{\mu}$ as a function, not a distribution here; fortunately since $\hat{\mu}$ is a distribution that is really a function in disguise, (i.e. $\hat{\mu}$ has a density function $g(x)$ such that $\langle \hat{\mu}, \varphi\rangle = \int \hat{\mu} g(x)\, dx$), we can recover $\hat{\mu}$ by applying it to an appropriate delta-function: $\hat{\mu}(\xi) = \langle \hat{\mu}, \delta(y-\xi)\rangle$. So following through with this plan and recalling that $\mathcal{F}\delta(y-\xi) = e^{-i\langle \xi, y\rangle}$,

$$\begin{aligned} \hat{\mu}_N(\xi) &= \left\langle \mu_N, e^{-i\langle \xi, y\rangle} \right\rangle = \lim_{y\to\infty} \frac{1}{y} \int_{\log 2}^{y} \exp\left(-i \left\langle \xi, \sum_{j=1}^{N} \mu_j(y) + b_0 \right\rangle\right) dy \\ &= \lim_{\substack{M_i\to\infty \\ 0\le i\le N}} \frac{1}{\prod M_i} \int_{\log 2}^{M_1} \int_{\log 2}^{M_2} \cdots \int_{\log 2}^{M_N} \exp\left(-i \left\langle \xi, \sum_{j=1}^{N} \mu_j(y_j) + b_0 \right\rangle\right) \mathbf{dy} \end{aligned}$$

But this integral is much simpler, because the variables separate:

$$\hat{\mu}_N(\xi) = \exp(-i\,\langle \xi, b_0\rangle) \prod_{j=1}^{N} \frac{1}{M_j} \int_{\log 2}^{M_j} \exp\Big(-i\,\langle \xi, \mu_j(y_j)\rangle\Big)\, dy_j$$

But each integral appearing on the inside is just the Fourier transform of a simpler function! So rewriting this more suggestively and taking the limit as $N \to \infty$, we see that

$$\hat{\mu}(\xi) = \lim_{N\to\infty} \exp i \Big(\sum_{m=1}^{r} c(q, a_m) \Big) \prod_{j=1}^{N} \hat{\mu}_{\gamma_j}(\xi)$$

where $\hat{\mu}_{\gamma_j}$ is the Fourier transform of a typical term

$$\mu_{\gamma_j} = -\Big(\frac{\bar{\chi}(a_1)e^{i\gamma_j y}}{\frac{1}{2}+i\gamma_j} + \frac{\chi(a_1)e^{-i\gamma_j y}}{\frac{1}{2}-i\gamma_j}, \dots, \frac{\bar{\chi}(a_r)e^{i\gamma_j y}}{\frac{1}{2}+i\gamma_j} + \frac{\chi(a_r)e^{-i\gamma_j y}}{\frac{1}{2}-i\gamma_j}\Big)$$

So we see what is really happening: $\hat{\mu}$ is the product of the Fourier transforms of "smaller" distributions due to each individual zero of the associated L-function, and a single term representing the Chebyshev bias. The $\hat{\mu}_\gamma$'s separate only thanks to GSH, and come out as a product.

Before we move on to the next step and compute $\hat{\mu}_\gamma$, we note that there is another way to understand why we end up with this product form for the Fourier Transform. We can illustrate this best in the case of only two roots, γ_1 and γ_2. When we are evaluating the big integral, all that matters is how much of the time the *sum* $\mu_1(y)+\mu_2(y)$ is equal to a given value, because we are essentially taking $\int f\big(\mu_1(y)+\mu_2(y)\big)$. Now if μ_1 and μ_2 were related to each other in some complicated way, we could't simplify this. But we know from probability theory that if $\mu_1(y)$ and $\mu_2(y)$ are independent variables, the probability that their sum is z is given by:

$$p(z) = \int_y \mu_1(y)\mu_2(z-y)\,dy = (\mu_1 * \mu_2)(z)$$

Where $(\mu_1 * \mu_2)(z)$ is of course the convolution of μ_1 with μ_2. Of course this isn't a probability question, it's an integral; but the value of the integral is indeed determined by the measure of the set on which $\mu_1(y)+\mu_2(y) = z$, and measures behave the same way as convolutions. GSH and the Weyl equidistribution theorem combine to assure us that the $\mu_j(y)$ are independent. The justification for taking such a circuitous argument is that the Fourier transform behaves very nicely under convolution, since $(f * g)\hat{}(y) = \hat{f}(y)\hat{g}(y)$. This convolution argument generalizes perfectly well to multiple dimensions; we just take

$$(\mu_1 * \mu_2 * \cdots * \mu_N)(z) = \int_{y_1}\int_{y_2}\cdots\int_{y_N} \mu_1(y_1)\mu_2(y_2)\cdots\mu_{N-1}(y_{N-1})\mu_N(z-\sum_{j=1}^{N-1} y_j)\,d\mathbf{y}$$

The same product formula for Fourier transforms applies in many variables, and we can again derive the product formula for $\hat{\mu}$; in fact this whole argument can be made rigorous using measure theory and the Weyl theorem to evaluate the integral as a convolution product.

Now to finish computing $\hat{\mu}$ we must evaluate $\hat{\mu}_\gamma$ for a typical γ. Recall that

$$\mu_\gamma = -\Big(\frac{\bar{\chi}(a_1)e^{i\gamma y}}{\frac{1}{2}+i\gamma} + \frac{\chi(a_1)e^{-i\gamma y}}{\frac{1}{2}-i\gamma}, \dots, \frac{\bar{\chi}(a_r)e^{i\gamma y}}{\frac{1}{2}+i\gamma} + \frac{\chi(a_r)e^{-i\gamma y}}{\frac{1}{2}-i\gamma}\Big)$$

Now let $\bar{\chi}(a_j) = u_j + iv_j$, so after some easy trigonometric substitutions

$$\mu_\gamma = -\frac{2}{R_\gamma}\left(u_1 \sin(\gamma y + w_\gamma) + v_1 \cos(\gamma y + w_\gamma), \ldots, u_r \sin(\gamma y + w_\gamma) + v_r \cos(\gamma y + w_\gamma)\right)$$

where we have taken $w_\gamma = \cot^{-1}(2\gamma)$ and $R_\gamma = \sqrt{1/4 + \gamma^2}$. So we once again apply the rule $\hat{\mu}_\gamma(\xi) = \int_{\mathbb{R}^r} \exp\big(i\,\langle \mu(y), \xi\rangle\big)\, d\mathbf{y}$ that we explained above. We can simplify this integral greatly by noting that $\sin(\gamma y)$ has density

$$\rho(t) = \begin{cases} \frac{1}{\pi\sqrt{1-t^2}} & \text{if } |t| < 1, \\ 0 & \text{otherwise} \end{cases}$$

That is just a fancy way of saying that the measure of the set on which $\sin(\gamma y) = t$ on the interval $[-M, M]$ is equal to $2M\rho(t)$ as $M \to \infty$. Of course, when $\sin(\gamma y) = t$, $\cos(\gamma y) = +\sqrt{1-t^2}$ half of the time, and $-\sqrt{1-t^2}$ the other half of the time. So thinking in terms of a Lebesque integral, we can replace $u_j \sin(\gamma y + w_\gamma) + v_j \cos(\gamma y + w_\gamma)$ with $u_j t \pm v_j \sqrt{1-t^2}$ and $d\mathbf{y}$ by $dt \big/ 2\pi\sqrt{(1-t^2)}$ to conclude that

$$\begin{aligned}\hat{\mu}_\gamma(\xi) =& \frac{1}{2}\int_{-1}^{1} \exp\left(iR_\gamma \sum_{m=1}^{r} \xi_m (u_m t + v_m \sqrt{1-t^2})\right) \frac{dt}{\pi\sqrt{(1-t^2)}} \\ &+ \frac{1}{2}\int_{-1}^{1} \exp\left(iR_\gamma \sum_{m=1}^{r} \xi_m (u_m t - v_m \sqrt{1-t^2})\right) \frac{dt}{\pi\sqrt{(1-t^2)}}\end{aligned}$$

Let $U = \sum_{m=1}^{r} \xi_m u_m$ and $V = \sum_{m=1}^{r} \xi_m v_m$; then

$$\begin{aligned}\hat{\mu}_\gamma(\xi) &= \frac{1}{\pi}\int_{-1}^{1} \frac{1}{2}\Big[\exp\left(iR_\gamma(Ut + V\sqrt{1-t^2})\right) + \exp\left(iR_\gamma(Ut - V\sqrt{1-t^2})\right)\Big] \frac{dt}{\sqrt{1-t^2}} \\ &= \frac{1}{\pi}\int_{-1}^{1} \exp(iR_\gamma U t)\cos(R_\gamma V\sqrt{1-t^2}) \frac{dt}{\sqrt{1-t^2}}\end{aligned}$$

But this this integral is just a Bessel function, so

$$\hat{\mu}_\gamma(\xi) = J_0\left(R_\gamma\sqrt{U^2 + V^2}\right)$$

where $J_0(z)$ is defined by

$$J_0(z) = \sum_{0}^{\infty} \frac{(-1)^m (\frac{1}{2}z)^{2m}}{(m!)^2}$$

So putting it all together, we have the following product formula for $\hat{\mu}$:

$$\hat{\mu}_{q;a_1,\ldots,a_r}(\xi) = \exp\left(i\sum_{j=1}^{r} c(q,a_j)\xi_j\right) \times \prod_{\substack{\chi\neq\chi_0 \\ \chi \bmod q}} \prod_{\gamma_\chi>0} J_0\left(\frac{2\left|\sum_{j=1}^{r}\chi(a_j)\xi_j\right|}{\sqrt{\frac{1}{4}+\gamma_\chi{}^2}}\right)$$

The factor of $\exp\left(i\sum_{j=1}^{r} c(q,a_j)\xi_j\right)$ arises from the Chebyshev bias, and the product is of course over zeroes $\frac{1}{2}+i\gamma_\chi$ of the L-function $L(x,\chi)$.
We can develop analagous product formulas for $\mu_{q;R,N}$, which measures the discrepancy between quadratic residues and nonresidues, and for μ_1, which measures $\pi(x)-\mathrm{Li}(x)$. Let $\pi_N(x,q)$ be the prime-counting function for quadratic nonresidues mod q, and define $\pi_R(x,q)$ analagously but for quadratic residues. Then $\mu_{q;R,N}$ is defined to be the limiting distribution of:

$$E_{q;R,N}(x) = (\pi_R(x,q) - \pi_N(x,q))\frac{\log x}{\sqrt{x}}$$

while μ_1 is defined to be the limiting distribution of

$$E_1 = (\pi(x) - \mathrm{Li}(x))\frac{\log x}{\sqrt{x}}$$

Then we can complete the derivation we just did for $\mu_{q;a_1,\ldots,a_r}$ word for word, with this exception; every time we used the characteristic function $\mathbf{1}_a(n) = \frac{1}{\varphi(q)}\sum_{\chi \bmod q}\bar{\chi}(a)\chi(n)$ we instead use the difference of the characteristic functions of the quadratic residues vs. nonresidues in computing $\mu_{q;R,N}$. That is, we take χ to be $\mathbf{1}_{\text{residues}} - \mathbf{1}_{\text{nonresidues}}$. But this is just a silly way of writing the quadratic character,

$$\chi_q(a) = \begin{cases} +1 & \text{if a is a quadratic residue,} \\ -1 & \text{if a is a quadratic nonresidue} \end{cases}$$

When we compute μ_1, the right character is the trivial character χ_0, and our L-function is just the Riemann Zeta function. So going through the arithmetic we find that

$$\hat{\mu}_{q;R,N}(\xi) = \exp(i\xi)\prod_{\gamma>0} J_0\left(\frac{2\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right)$$

where the product runs over zeroes $\frac{1}{2}+i\gamma$ of $L(s,\chi_q)$. The formula for $\hat{\mu}_1(\xi)$ is identical, except it runs over zeroes of $\zeta(s)$.

Applications of the Product Formula

In this section we explore some of the interesting results we can establish with the aid of the product formulas for the various $\hat{\mu}$'s. We start with a quick demonstration that the logarithmic density of the quadratic non-residues is always strictly larger than one half. Recall that

$$\delta(P_{q;R,N}) = \int_0^\infty d\mu_{q;R,N}(t)$$

Now the Bessel function J_0 is even, so $\hat{\mu}_{q;R,N}(\xi) = e^{i\xi} f(\xi)$ for the even function $f(\xi) = \prod_\gamma J_0\left(2\xi/\sqrt{\frac{1}{4}+\gamma^2}\right)$. But a basic fact about Fourier transforms is that if $f_\beta(x) = f(x+\beta)$, then $\hat{f}_\beta(y) = e^{i\beta y}\hat{f}(y)$. So it is clear that $\mu_{q;R,N}$ is symmetric about $t=-1$, so

$$\delta(P_{q;R,N}) = \int_{-1}^\infty d\mu_{q;R,N}(t) - \int_{-1}^0 d\mu_{q;R,N}(t) = \frac{1}{2} - \int_{-1}^0 d\mu_{q;R,N}(t)$$

The integral we're subtracting from $1/2$ is strictly larger than zero because μ is entire and hence cannot be identically zero on $(-1,0)$.
A more interesting fact is that in the limit as $q\to\infty$, the Chebyshev bias dissolves and $\mu_{q;R,N}$ tends to a Gaussian. The idea is simple: we expand $J_0(z)$ out in power series, $J_0(z) = 1+\frac{1}{4}z^2+O(z^4)$, and note that the error in replacing $\log(1-z^2+O(z^4)$ is certainly no worse than $O(z^4))$, to conclude that

$$\log\hat{\mu}_{q;R,N}(z) = iz + \sum_{\gamma>0}\log\left(1-\frac{1}{4}\frac{4z^2}{\frac{1}{4}+\gamma^2}\right) + O\left(\sum_{\gamma>0}\frac{z^4}{\left(\frac{1}{4}+\gamma^2\right)^2}\right)$$

So taking $z = \xi\,/\log q$ we find that when $|\xi|<A$,

$$\log\hat{\mu}_{q;R,N}\left(\frac{\xi}{\log q}\right) = \frac{i\xi}{\sqrt{\log q}} - \frac{\xi^2}{\log q}\sum_{\gamma>0}\frac{1}{\frac{1}{4}+\gamma^2} + O\left(\frac{A^4}{\log^2 q}\sum_{\gamma>0}\frac{1}{\left(\frac{1}{4}+\gamma^2\right)^2}\right)$$

A well known result following from the partial fraction expansion of $L(x,\chi)$, is the formula

$$\sum_{\gamma>0}\frac{1}{\frac{1}{4}+\gamma^2} = \frac{1}{2}\log\left(\frac{q}{\pi}\right) - \frac{1}{2}\gamma - \frac{1}{2}(\chi(-1)+1)\log 2 + \frac{L'}{L}(1,\chi)$$

which can be obtained by logarithmically differentiating the infinite product form for

$$\xi(s) = e^{A+Bs} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{s/\rho}$$

Now we apply the useful fact that under GRH, $\frac{L'}{L}(1,\chi) = O(\log\log q)$, so

$$\sum_{\gamma>0} \frac{1}{\frac{1}{4}+\gamma^2} = \frac{1}{2}\log q + O(\log\log q)$$

Both of these results are in [Davenport]. Furthermore, $\left(\frac{1}{4}+\gamma^2\right)^2 \geq \frac{1}{4}\left(\frac{1}{4}+\gamma^2\right)$, so

$$\sum_{\gamma>0} \frac{1}{\left(\frac{1}{4}+\gamma^2\right)^2} \leq 4 \sum_{\gamma>0} \frac{1}{\frac{1}{4}+\gamma^2}$$

So putting these bounds together we find that when $|\xi| < A$

$$\log \hat{\mu}_{q;R,N}\left(\frac{\xi}{\sqrt{\log q}}\right) = -\frac{1}{2}\xi^2 + O\left(\frac{A}{\sqrt{\log q}} + \frac{A^2 \log\log q}{\log q} + \frac{A^4}{\log q}\right)$$

and $\hat{\mu}_{q;R,N}(\xi/\log q)$ tends uniformly to the Gaussian $e^{-\xi^2}$ on $[-A, A]$ as $q \to \infty$. Thus we apply an analytic result called Lévy's criterion [Lévy] to conclude that the measures $\hat{\mu}_{q;R,N}$ converge in measure to the Gaussian. Applying Fourier inversion, we conclude that $\mu_{q;R,N}(y) \to \sqrt{q}\exp(qy^2/2)$. As a corollary, we obtain the result that $\delta(P_{q;R,N}) \to \frac{1}{2}$ as $q \to \infty$.
We can use similar reasoning to obtain the more general result that for any choice of $(a_1, \ldots, a_r)$ with r fixed, the distribution $\hat{\mu}_{q;a_1,\ldots,a_r}$ tends to a Gaussian in measure as $q \to \infty$, and the corresponding logarithmic density is thus $\delta(P_{q;a_1,\ldots,a_r}) = 1/r!$. The proof isn't different enough to be worth going through. And intuitively, it is clear that since the largest effect causing bias—the quadratic vs. non-quadratic discrepancy—dissolves as q becomes sufficiently large, then any other biases will also disappear as $q \to \infty$.
In spite of the fact that the bias tends to zero as q becomes large, we show now that it is equal to zero only in certain very limited cases. In fact, we will show that $\mu_{q;a_1,\ldots,a_r}$ is symmetric in $(x_1, \ldots, x_r)$ if and only if either

1. $r = 2$ and $c(q, a_1) = c(q, a_2)$, or

2. $r = 3$ and there exists $\rho \neq 1$ such that ρ satisfies the congruences modulo q

$$\rho^3 \equiv 1, \quad a_2 \equiv a_1\rho, \quad a_3 \equiv a_1\rho^2$$

Now we can see that the factor of $\exp\left(i\sum_{j=1}^{r} c(q,a_j)\xi_j\right)$ shifts the mean of μ to $-\big(c(q,a_1),\dots,c(q,a_r)\big)$, so the equality of the $c(q,a_j)$'s is certainly a necessary condition. Thus the question of the symmetry of μ devolves upon the product of Bessel functions. Our strategy is to show that in the cases we have just described, the Bessel function product is necessarily symmetric in the x_j because its argument is symmetric; and we will then show conversely that this is a necessary condition. The argument of the Bessel functions is $2\left|\sum_{j=1}^{r}\chi(a_j)\xi_j\right| \Big/ \sqrt{\frac{1}{4}+\gamma_\chi{}^2}$, so we define

$$B_\chi(\xi_1,\dots,\xi_r) = \left|\sum_{j=1}^{r}\chi(a_j)\xi_j\right|$$

and prove the following

Lemma: *$B_\chi(\xi_1,\dots,\xi_r)$ is symmetric in $(\xi_1,\dots,\xi_r)$ for all χ if and only if one of the two conditions listed above is satisfied*

Proof:

If $r=2$ we have $B_\chi(\xi_1,\xi_2) = |\chi(a_1\xi_1+\chi(a_2)\xi_2|$. But $|\chi(a_1)| = |\chi(a_2)| = 1$, so

$$\begin{aligned}|\chi(a_1)\xi_1+\chi(a_2)\xi_2| &= \left|\overline{\chi(a_1)\chi(a_2)}(\chi(a_1)\xi_1+\chi(a_2)\xi_2)\right| = \left|\overline{\chi(a_2)}\xi_1+\overline{\chi(a_1)}\xi_2\right| \\ &= \left|\overline{\overline{\chi(a_2)}\xi_1+\overline{\chi(a_1)}\xi_2}\right| = |\chi(a_2)\xi_1+\chi(a_1)\xi_2| = B_\chi(\xi_2,\xi_1)\end{aligned}$$

If $r=3$ and a ρ exists as described, then

$$\chi(a_2)=\chi(a_1)\chi(\rho),\quad \chi(a_3)=\chi(a_1)\chi^2(\rho),\quad \chi^3(\rho)=1$$

Thus if we set $\tau=\chi(\rho)$, $\tau^3=1$ and we have $|\chi(a_1)\xi_1+\chi(a_2)\xi_2+\chi(a_3)\xi_3| = \left|\chi(a_1)\xi_1+\tau\chi(a_1)\xi_2+\tau^2\chi(a_1)\xi_3\right| = \left|\xi_1+\tau\xi_2+\tau^2\xi_3\right|$. But

$$\left|\xi_1+\tau\xi_2+\tau^2\xi_3\right| = \left|\tau(\xi_1+\tau\xi_2+\tau^2\xi_3)\right| = \left|\xi_3+\tau\xi_1+\tau^2\xi_2\right| = \left|\xi_2+\tau\xi_3+\tau^2\xi_1\right|$$

Furthermore, $\left|\xi_1+\tau\xi_2+\tau^2\xi_3\right| = \left|\overline{\xi_1+\tau\xi_2+\tau^2\xi_3}\right| = \left|\xi_1+\tau\xi_3+\tau^2\xi_2\right|$. So we conclude that $B_\chi(\xi_1,\xi_2,\xi_3)$ is symmetric in (ξ_1,ξ_2,ξ_3). Conversely, if $|\chi(a_1)\xi_1+\chi(a_2)\xi_2+\chi(a_3)\xi_3|$ is symmetric in (ξ_1,ξ_2,ξ_3) for all χ, then so is $|\xi_1+\chi(a_2/a_1)\xi_2+\chi(a_3/a_1)\xi_3|$. Hence $\operatorname{Re}\chi(a_2/a_1) = \operatorname{Re}\chi(a_3/a_1)$, and $\operatorname{Re}\chi(a_1/a_2) = \operatorname{Re}\chi(a_3/a_2)$. So if we let $w=\chi(a_2/a_1)$, then $\chi(a_3/a_1)=w^2$ and $w^3=1$. Since this is true for all χ, there must be a ρ fulfilling the desired congruences. An adaptation of this argument shows that the $B_\chi(\xi_1,\dots,\xi_r)$ cannot be symmetric when $r\geq 4$. For if it were, then any three of the a_i's would be related as they were above, and they could not then all be distinct. □

With this lemma, the rest of the proof essentially amounts to showing that if B_χ is not symmetric, then there is at least one permutation σ of $(\xi_1, \ldots, \xi_r)$ that will kill the symmetry in the product. This is neither difficult nor interesting, so we omit it here.

Our final application of the product formula is the most interesting one. Using it, we will show how one can numerically compute the logarithmic densities $\delta(P_1)$ and $\delta(P_{q;N,R})$ for $q = 3, 4, 5, 7, 11$, and 13. Recall that $P_1 = \{x \geq 2 \colon \pi(x) > \mathrm{Li}(x)\}$. Let $f_{q;N,R}(t)$ and $f_1(t)$ be the densities of $\mu_{q;R,N}$ and μ_1. It is more convenient to work with symmetric functions, so let ω be the the distribution whose density is $g(t) = f(t-1)$. Then its Fourier transform $\hat{\omega}(\xi) = \prod_{\gamma>0} J_0\left(2\xi \Big/ \sqrt{1/4+\gamma^2}\right)$ is symmetric about zero. We want to evaluate $\delta(P) = \int_{-\infty}^{1} d\omega$ for the various ω's. Now by symmetry, we have

$$\delta(P) = \frac{1}{2} + \frac{1}{2}\int_{-1}^{1} d\omega(t) = \frac{1}{2} + \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin u}{u}\hat{\omega}(u)\, du$$

To get the last equality, we rewrote the integral as $\int_{-\infty}^{\infty} \chi_{[-1,1]}\, d\omega(t)$ and then made use of Parseval's identity that $\langle f, g\rangle = \left\langle \hat{f}, \hat{g}\right\rangle$; the $\sin u/u$ term is the Fourier transform of the characteristic function $\chi_{[-1,1]}$.

The first step in evaluating this integral is to replace it cleverly with a finite sum. Apply the Poisson summation formula to the functions

$$\varphi(u) = \frac{1}{2\pi}\frac{\sin u}{u}\hat{\omega}(u) \text{ and } \hat{\varphi}(x) = \frac{1}{2}(\chi_{[-1,1]} * g)(x) = \frac{1}{2}\int_{x-1}^{x+1} d\omega(u)$$

Poisson summation is permissible here because the Bessel function is sufficiently well behaved that φ is is rapidly decreasing. So

$$\varepsilon \sum_{n\in\mathbb{Z}} \varphi(\varepsilon n) = \sum_{n\in\mathbb{Z}} \hat{\varphi}\left(\frac{n}{\varepsilon}\right) = \hat{\varphi}(0) + \sum_{n\neq 0} \hat{\varphi}\left(\frac{n}{\varepsilon}\right)$$

Of course the integral we want is equal to $\int \varphi(u)\, du = \hat{\varphi}(0)$, so we find that

$$\frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin u}{u}\hat{\omega}(u)\, du = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} \varepsilon\frac{\sin \varepsilon n}{\varepsilon n}\hat{\omega}(\varepsilon n) - \sum_{n\neq 0} \hat{\varphi}\frac{n}{\varepsilon}$$

Our plan is to replace the integral with the left-hand Riemann sum, and to estimate the error thus incurred. This is done by using the bound:

$$\omega\left[2\sum_{0<\gamma\leq T} R_\gamma, \infty\right) \leq \exp\left(-\frac{3}{4}\frac{(\sum_{0<\gamma\leq T} R_\gamma)^2}{\sum_{\gamma>T} {R_\gamma}^2}\right)$$

found in [Montgomery], where R_γ is again $2 \Big/ \sqrt{\frac{1}{4}+\gamma^2}$. This bound holds as long as the sums are nonempty. To get a good numerical bound, we look at the zeroes of the L-functions we're interested in and note that they all have $|\gamma| > 2$. So for any $\lambda \geq 0$, we can find T such that $0 \leq \lambda - 2 \sum\limits_{0<\gamma\leq T} T_\gamma < 2$, so our bound becomes

$$\omega[\lambda, \infty) < \exp\left(-\frac{3}{4}\frac{\left(\frac{1}{2}(\lambda-2)\right)^2}{\sum_{\gamma>T} R_\gamma^2}\right) \leq \exp\left(-\frac{3}{4}\frac{\left(\frac{1}{2}(\lambda-2)\right)^2}{\sum_{\gamma>0} R_\gamma^2}\right)$$

So using bounds obtained by looking at the appropriate sets of zeroes, we find that $\left(\sum_{\gamma>0} R_\gamma^2\right)^{-1} > .98$ for all the q's under consideration. So after a few more easy computations we find that the error in taking $\varepsilon = 1/20$ is at most 10^{-20}.

The next step is to replace the infinite sum $\sum_n \varepsilon \frac{\sin \varepsilon n}{\varepsilon n} \hat{\omega}(\varepsilon n)$ with a sum from $-C$ to C for some aptly chosen constant C. This is a routine computation that essentially amounts to bounding $\hat{\omega}$, and only involves the elementary estimate $|J_0(z)| \leq \min\left(1, \sqrt{2/(\pi\,|x|)}\right)$. What is significant is that in order to achieve 8-digit accuracy for $\delta(P_1)$ we only require $C = 50$, while $C = 25$ suffices for 4-digit accuracy on the other densities. So in practice these sums are quite reasonable to perform as they all have at most $100 \times 20 = 2000$ terms. Unfortunately, we are still left with a messy infinite product for $\hat{\omega}(u)$. We work around that by replacing it with the product of only those γ's less than some T times a polynomial $p(u)$ that we choose to approximate the remaining terms. That is, we take

$$\hat{\omega}(u) = p(u) \prod_{\gamma \leq T} J_0\left(\frac{2u}{\sqrt{\frac{1}{4}+\gamma^2}}\right)$$

where $p(u) = \sum\limits_{m=0}^{A} b_m u^{2m}$ approximates $\prod\limits_{\gamma>T} J_0(R_\gamma u) = \sum\limits_{m=0}^{\infty} b_m u^{2m}$. We can see that this expansion is valid by looking at the expansion for $J_0(z)$ and recalling that the sum $\sum\limits_{\gamma>T} \left(\frac{1}{4}+\gamma^2\right)^{-1}$ converges, say to some $M = M(T)$. From the power series for $J_0(z) = 1 + z^2/4 + \cdots$, we can see that $|b_m|$ will be less than the coefficients of

$$\prod_{\gamma>T} J_0\left(\frac{2u}{\sqrt{\frac{1}{4}+\gamma^2}}\right) \leq \prod_{\gamma>T} \exp\left(\frac{1}{4}\left(\frac{2u}{\sqrt{\frac{1}{4}+\gamma^2}}\right)^2\right) = \exp\left(u^2 \sum_{\gamma>T} \frac{1}{\frac{1}{4}+\gamma^2}\right)$$

In particular, this tells us that $|b_m| < M^m/m!$, so

$$\left|\sum_{m=A+1}^{\infty} b_m u^{2m}\right| < \sum_{A+1}^{\infty} \frac{M^m}{m!}|u|^{2m} < \frac{(Mu^2)^{A+1}}{(A+1)!}(1 + Mu^2 + (Mu^2)^2 + \cdots)$$

This last expression is a convergent geometric series equal to

$$\frac{(Mu^2)^{A+1}}{(A+1)!}\frac{1}{1-Mu^2}$$

if $Mu^2 < 1$, and in particular it is less than $2(Mu^2)^{A+1}/(A+1)!$ if $Mu^2 < \frac{1}{2}$. At this point we've done most of the work. To estimate the error, we plug this last estimate into the sum from $-C$ to C and then pick T large enough so we only need a small A. In fact, by using the zeroes with $\gamma < 10{,}000$ we only need to take $A = 1$ to calculate the densities of $P_{q;N,R}$. We certainly don't need to worry about Mu^2 exceeding $\frac{1}{2}$, since in practice M is very small for the T's under consideration. We can get b_1 easily enough by noting that

$$b_1 = -M(T) = -\left(\sum_{\gamma>0}\frac{1}{\frac{1}{4}+\gamma^2} - \sum_{0<\gamma<T}\frac{1}{\frac{1}{4}+\gamma^2}\right)$$

But the formula we've already seen from [Davenport] tells us that

$$\sum_{\gamma>0}\frac{1}{\frac{1}{4}+\gamma^2} = \frac{1}{2}\log\left(\frac{q}{\pi}\right) - \frac{1}{2}\gamma - \frac{1}{2}(\chi(-1)+1)\log 2 + \frac{L'}{L}(1,\chi)$$

(of course the γ on the right is Euler's constant).
We can evaluate L' by using Dirichlet's formula

$$\Gamma(s)L(s,\chi_q) = \int_0^{\infty}\frac{h(e^{-u})}{1-e^{uq}}u^{s-1}e^{-u}\,du$$

where $h(x) = \sum_1^{q-1}\chi_q(m)x^{m-1}$, and then differentiating it at $s = 1$:

$$L'(1,\chi_q) = \gamma L(1,\chi_q) + \int_0^{\infty}\frac{h(e^{-u})}{1-e^{-uq}}\log(u)e^{-u}\,du$$

This integral can be performed numerically. And in class we developed a number of easy ways to compute $L(1,\chi)$ given the zeroes of the L-function with imaginary part $< T$. The procedure used for P_1 is similar, except that we need to get both b_1 and b_2, which can be done by brute force on the zeroes of $\zeta(s)$ with $|\gamma| < 100{,}000$. The results, which are extremely interesting (provided that they're true!) are as follows:

$$\begin{aligned}
\delta(P_1) &= 0.99999973\ldots \\
\delta(P_{3;N,R}) &= 0.9990\ldots \\
\delta(P_{4;N,R}) &= 0.9959\ldots \\
\delta(P_{5;N,R}) &= 0.9954\ldots \\
\delta(P_{7;N,R}) &= 0.9782\ldots \\
\delta(P_{11;N,R}) &= 0.9167\ldots \\
\delta(P_{13;N,R}) &= 0.9443\ldots
\end{aligned}$$

References

[Davenport] H. Davenport, *Multiplicative Number Theory* (2nd ed), Graduate Texts in Mathematics **74**, Springer, Berlin, 1980

[Montgomery] H.L. Montgomery, "The Zeta Functions and Prime Numbers," pp. 14-24 in *Proceedings of the Queen's Number Theory Conference, 1979*, (edited by P. Ribenboim), Queen's University, Kingston (Ont.), 1980

[Lévy] P. Lévy, "Sur la détermination des lois de probabilité par leurs fonctions caractéristiques," *C. R. Acad. Sci. Paris* **175** (1922)

[R&S] M. Rubinstein and P. Sarnak, "Chebyshev's Bias," *Experimental Mathematics, Vol. 3 (1994)*