Unexpected biases in the distribution of consecutive primes

Robert J. Lemke Oliver, Kannan Soundararajan
Stanford University

Chebyshev's Bias

Let $\pi(x; q, a) := \#\{p < x : p \equiv a (\text{mod } q)\}$.

Chebyshev's Bias

Let $\pi(x; q, a) := \#\{p < x : p \equiv a (\text{mod } q)\}$. Under GRH, we have

$$\pi(x; q, a) = \frac{\mathrm{li}(x)}{\phi(q)} + O(x^{1/2+\epsilon}) \quad \text{if } (a, q) = 1.$$

Chebyshev's Bias

Let $\pi(x; q, a) := \#\{p < x : p \equiv a (\text{mod } q)\}$. Under GRH, we have

$$\pi(x; q, a) = \frac{\mathrm{li}(x)}{\phi(q)} + O(x^{1/2+\epsilon}) \quad \text{if } (a, q) = 1.$$

Observation (Chebyshev)

Quadratic residues seem slightly less frequent than non-residues.

Chebyshev's Bias

Let $\pi(x; q, a) := \#\{p < x : p \equiv a (\text{mod } q)\}$. Under GRH, we have

$$\pi(x; q, a) = \frac{\mathrm{li}(x)}{\phi(q)} + O(x^{1/2+\epsilon}) \quad \text{if } (a, q) = 1.$$

Observation (Chebyshev)

Quadratic residues seem slightly less frequent than non-residues.

Theorem (Rubinstein-Sarnak)

Under GRH$(+\epsilon)$, $\pi(x; 3, 2) > \pi(x; 3, 1)$ *for* 99.9% *of* x,

Chebyshev's Bias

Let $\pi(x; q, a) := \#\{p < x : p \equiv a(\text{mod } q)\}$. Under GRH, we have

$$\pi(x; q, a) = \frac{\mathrm{li}(x)}{\phi(q)} + O(x^{1/2+\epsilon}) \quad \text{if } (a, q) = 1.$$

Observation (Chebyshev)

Quadratic residues seem slightly less frequent than non-residues.

Theorem (Rubinstein-Sarnak)

Under GRH$(+\epsilon)$, $\pi(x; 3, 2) > \pi(x; 3, 1)$ *for* 99.9% *of* x, *and analogous results hold for any* q.

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$,

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\text{mod } q) \text{ for } 0 \leq i < r\}.$$

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\text{mod } q) \text{ for } 0 \leq i < r\}.$$

We expect that $\pi(x; q, \mathbf{a}) \sim \mathrm{li}(x)/\phi(q)^r$,

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\text{mod } q) \text{ for } 0 \leq i < r\}.$$

We expect that $\pi(x; q, \mathbf{a}) \sim \mathrm{li}(x)/\phi(q)^r$, but little is known:

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\text{mod } q) \text{ for } 0 \leq i < r\}.$$

We expect that $\pi(x; q, \mathbf{a}) \sim \mathrm{li}(x)/\phi(q)^r$, but little is known:

- If $r = 2$ and $\phi(q) = 2$, then each $\mathbf{a}$ occurs infinitely often

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\text{mod } q) \text{ for } 0 \leq i < r\}.$$

We expect that $\pi(x; q, \mathbf{a}) \sim \mathrm{li}(x)/\phi(q)^r$, but little is known:

- If $r = 2$ and $\phi(q) = 2$, then each $\mathbf{a}$ occurs infinitely often
- Shiu: The pattern $(a, a, \ldots, a)$ occurs infinitely often

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\bmod\ q) \text{ for } 0 \leq i < r\}.$$

We expect that $\pi(x; q, \mathbf{a}) \sim \mathrm{li}(x)/\phi(q)^r$, but little is known:

- If $r = 2$ and $\phi(q) = 2$, then each $\mathbf{a}$ occurs infinitely often
- Shiu: The pattern $(a, a, \ldots, a)$ occurs infinitely often
- Maynard: $\pi(x; q, (a, a, \ldots, a)) \gg \pi(x)$

Patterns of consecutive primes

Let $r \geq 1$ and $\mathbf{a} = (a_1, \ldots, a_r)$, and set

$$\pi(x; q, \mathbf{a}) := \#\{p_n < x : p_{n+i} \equiv a_{i+1} (\text{mod } q) \text{ for } 0 \leq i < r\}.$$

We expect that $\pi(x; q, \mathbf{a}) \sim \mathrm{li}(x)/\phi(q)^r$, but little is known:

- If $r = 2$ and $\phi(q) = 2$, then each $\mathbf{a}$ occurs infinitely often
- Shiu: The pattern $(a, a, \ldots, a)$ occurs infinitely often
- Maynard: $\pi(x; q, (a, a, \ldots, a)) \gg \pi(x)$

Question

Are there biases between the different patterns $\mathbf{a} (\text{mod } q)$?

The primes (mod 10)

Let $\pi(x_0) = 10^7$.

The primes (mod 10)

Let $\pi(x_0) = 10^7$. We find:

a	$\pi(x_0; 10, a)$
1	2,499,755
3	2,500,209

a	$\pi(x_0; 10, a)$
7	2,500,283
9	2,499,751

The primes (mod 10)

Let $\pi(x_0) = 10^7$. We find:

a	b	$\pi(x_0; 10, (a, b))$
1	1	2,499,755
	3	
	7	
	9	
	3	2,500,209

a	$\pi(x_0; 10, a)$
7	2,500,283
9	2,499,751

The primes (mod 10)

Let $\pi(x_0) = 10^7$. We find:

a	b	$\pi(x_0; 10, (a, b))$
1	1	446,808
	3	756,071
	7	769,923
	9	526,953
	3	2,500,209

a	$\pi(x_0; 10, a)$
7	2,500,283
9	2,499,751

The primes (mod 10)

Let $\pi(x_0) = 10^7$. We find:

a	b	$\pi(x_0; 10, (a, b))$
1	1	446,808
	3	756,071
	7	769,923
	9	526,953
3	1	593,195
	3	422,302
	7	714,795
	9	769,915

a	b	$\pi(x_0; 10, (a, b))$
7	1	639,384
	3	681,759
	7	422,289
	9	756,851
9	1	820,368
	3	640,076
	7	593,275
	9	446,032

The primes (mod 10)

Let $\pi(x_1) = 10^8$. We find:

a	b	$\pi(x_1; 10, (a, b))$
1	1	4,623,041
	3	7,429,438
	7	7,504,612
	9	5,442,344
3	1	6,010,981
	3	4,442,561
	7	7,043,695
	9	7,502,896

a	b	$\pi(x_1; 10, (a, b))$
7	1	6,373,982
	3	6,755,195
	7	4,439,355
	9	7,431,870
9	1	7,991,431
	3	6,372,940
	7	6,012,739
	9	4,622,916

The primes (mod 3)

Let $\pi(x_0) = 10^7$.

The primes (mod 3)

Let $\pi(x_0) = 10^7$. We have

$$\pi(x_0; 3, 1) = 4{,}999{,}505 \quad \text{and} \quad \pi(x_0; 3, 2) = 5{,}000{,}494$$

The primes (mod 3)

Let $\pi(x_0) = 10^7$. We have

$$\pi(x_0; 3, 1) = 4{,}999{,}505 \quad \text{and} \quad \pi(x_0; 3, 2) = 5{,}000{,}494$$

while for $r \geq 2$, we have

a	$\pi(x_0; 3, \mathbf{a})$
1,1	
1,2	
2,1	
2,2	

The primes (mod 3)

Let $\pi(x_0) = 10^7$. We have

$$\pi(x_0; 3, 1) = 4{,}999{,}505 \quad \text{and} \quad \pi(x_0; 3, 2) = 5{,}000{,}494$$

while for $r \geq 2$, we have

a	$\pi(x_0; 3, \mathbf{a})$
1,1	2,203,294
1,2	2,796,209
2,1	2,796,210
2,2	2,204,284

The primes (mod 3)

Let $\pi(x_0) = 10^7$. We have

$$\pi(x_0; 3, 1) = 4{,}999{,}505 \quad \text{and} \quad \pi(x_0; 3, 2) = 5{,}000{,}494$$

while for $r \geq 2$, we have

a	$\pi(x_0; 3, \mathbf{a})$
1,1	2,203,294
1,2	2,796,209
2,1	2,796,210
2,2	2,204,284

and

a	$\pi(x_0; 3, \mathbf{a})$
1,1,1	928,276
1,1,2	1,275,018
1,2,1	1,521,062
1,2,2	1,275,147
2,1,1	1,275,018
2,1,2	1,521,191
2,2,1	1,275,147
2,2,2	929,137

The primes (mod 3)

Let $\pi(x_0) = 10^7$. We have

$$\pi(x_0; 3, 1) = 4{,}999{,}505 \quad \text{and} \quad \pi(x_0; 3, 2) = 5{,}000{,}494$$

while for $r \geq 2$, we have

a	$\pi(x_0; 3, \mathbf{a})$
1,1	2,203,294
1,2	2,796,209
2,1	2,796,210
2,2	2,204,284

and

a	$\pi(x_0; 3, \mathbf{a})$
1,1,1	928,276
1,1,2	1,275,018
1,2,1	1,521,062
1,2,2	1,275,147
2,1,1	1,275,018
2,1,2	1,521,191
2,2,1	1,275,147
2,2,2	929,137

Observation

The primes dislike to repeat themselves (mod q).

What's the deal?

What's the deal?

We conjecture that:

- There are large secondary terms in the asymptotic for $\pi(x; q, \mathbf{a})$

What's the deal?

We conjecture that:

- There are large secondary terms in the asymptotic for $\pi(x; q, \mathbf{a})$
- The dominant factor is the number of $a_i \equiv a_{i+1} (\text{mod } q)$

What's the deal?

We conjecture that:

- There are large secondary terms in the asymptotic for $\pi(x; q, \mathbf{a})$
- The dominant factor is the number of $a_i \equiv a_{i+1} \pmod{q}$
- There are smaller, somewhat erratic factors that affect non-diagonal $\mathbf{a}$

The conjecture: explicit version

Conjecture (LO & S)

Let $\mathbf{a} = (a_1, \ldots, a_r)$ *with* $r \geq 2$.

The conjecture: explicit version

Conjecture (LO & S)

Let $\mathbf{a} = (a_1, \ldots, a_r)$ *with* $r \geq 2$. *Then*

$$\pi(x; q, \mathbf{a}) = \frac{\mathrm{li}(x)}{\phi(q)^r} \left[1 + c_1(q; \mathbf{a}) \frac{\log \log x}{\log x} + \frac{c_2(q; \mathbf{a})}{\log x} + O\left(\log^{-7/4} x \right) \right],$$

The conjecture: explicit version

Conjecture (LO & S)

Let $\mathbf{a} = (a_1, \ldots, a_r)$ *with* $r \geq 2$*. Then*

$$\pi(x; q, \mathbf{a}) = \frac{\mathrm{li}(x)}{\phi(q)^r} \left[1 + c_1(q; \mathbf{a}) \frac{\log\log x}{\log x} + \frac{c_2(q; \mathbf{a})}{\log x} + O\left(\log^{-7/4} x\right)\right],$$

where

$$c_1(q; \mathbf{a}) = \frac{\phi(q)}{2} \left(\frac{r-1}{\phi(q)} - \#\{1 \leq i < r : a_i \equiv a_{i+1} (mod\ q)\}\right),$$

The conjecture: explicit version

Conjecture (LO & S)

Let $\mathbf{a} = (a_1, \ldots, a_r)$ *with* $r \geq 2$. *Then*

$$\pi(x; q, \mathbf{a}) = \frac{\mathrm{li}(x)}{\phi(q)^r} \left[1 + c_1(q; \mathbf{a}) \frac{\log \log x}{\log x} + \frac{c_2(q; \mathbf{a})}{\log x} + O\left(\log^{-7/4} x \right) \right],$$

where

$$c_1(q; \mathbf{a}) = \frac{\phi(q)}{2} \left(\frac{r-1}{\phi(q)} - \#\{1 \leq i < r : a_i \equiv a_{i+1} (mod\ q)\} \right),$$

and $c_2(q; \mathbf{a})$ *is complicated but explicit.*

An example

Example

Let $q = 3$ or 4.

An example

Example

Let $q = 3$ or 4. Then

$$\pi(x;q,(a,b)) = \frac{\mathrm{li}(x)}{4}\left[1 \pm \left(\frac{\log\log x}{2\log x} + \frac{\log 2\pi/q}{2\log x}\right)\right] + O\left(\frac{x}{\log^{11/4} x}\right).$$

An example

Example

Let $q = 3$ or 4. Then

$$\pi(x; q, (a, b)) = \frac{\mathrm{li}(x)}{4}\left[1 \pm \left(\frac{\log\log x}{2\log x} + \frac{\log 2\pi/q}{2\log x}\right)\right] + O\left(\frac{x}{\log^{11/4} x}\right).$$

Conjecture (LO & S)

Let $q = 3$ or 4. If $a \not\equiv b (mod\ q)$, then for all $x > 5$,

$$\pi(x; q, (a, b)) > \pi(x; q, (a, a)).$$

Comparison with numerics: $q = 3$

	x	$\pi(x;3,(1,1))$	$\pi(x;3,(1,2))$
Actual	10^9	$1.132 \cdot 10^7$	$1.411 \cdot 10^7$
Conj.		$1.156 \cdot 10^7$	$1.387 \cdot 10^7$

Comparison with numerics: $q = 3$

	x	$\pi(x;3,(1,1))$	$\pi(x;3,(1,2))$
Actual	10^9	$1.132 \cdot 10^7$	$1.411 \cdot 10^7$
Pred.		$1.137 \cdot 10^7$	$1.405 \cdot 10^7$
Conj.		$1.156 \cdot 10^7$	$1.387 \cdot 10^7$

Comparison with numerics: $q = 3$

	x	$\pi(x;3,(1,1))$	$\pi(x;3,(1,2))$
Actual	10^9	$1.132 \cdot 10^7$	$1.411 \cdot 10^7$
Pred.		$1.137 \cdot 10^7$	$1.405 \cdot 10^7$
Conj.		$1.156 \cdot 10^7$	$1.387 \cdot 10^7$
	10^{10}	$1.024 \cdot 10^8$	$1.251 \cdot 10^8$
		$1.028 \cdot 10^8$	$1.247 \cdot 10^8$
		$1.042 \cdot 10^8$	$1.233 \cdot 10^8$
	10^{11}	$9.347 \cdot 10^8$	$1.124 \cdot 10^9$
		$9.383 \cdot 10^8$	$1.121 \cdot 10^9$
		$9.488 \cdot 10^8$	$1.110 \cdot 10^9$
	10^{12}	$8.600 \cdot 10^9$	$1.020 \cdot 10^{10}$
		$8.630 \cdot 10^9$	$1.017 \cdot 10^{10}$
		$8.712 \cdot 10^9$	$1.009 \cdot 10^{10}$

The conjecture when $q = 5$

When $q = 5$,

The conjecture when $q = 5$

When $q = 5$, we predict that

$$c_2(5; (a, a)) = \frac{-3 \log(2\pi/5)}{2},$$

The conjecture when $q = 5$

When $q = 5$, we predict that

$$c_2(5; (a, a)) = \frac{-3 \log(2\pi/5)}{2},$$

and if $a \neq b$, then $c_2(5; (a, b)) =$

$$\frac{\log(2\pi/5)}{2} + \frac{5}{2}\Re\left(L(0,\chi)L(1,\chi)A_{5,\chi}\left[\bar{\chi}(b-a) + \frac{\bar{\chi}(b) - \bar{\chi}(a)}{4}\right]\right),$$

The conjecture when $q = 5$

When $q = 5$, we predict that

$$c_2(5; (a, a)) = \frac{-3 \log(2\pi/5)}{2},$$

and if $a \neq b$, then $c_2(5; (a, b)) =$

$$\frac{\log(2\pi/5)}{2} + \frac{5}{2}\Re\left(L(0,\chi)L(1,\chi)A_{5,\chi}\left[\bar{\chi}(b-a) + \frac{\bar{\chi}(b) - \bar{\chi}(a)}{4}\right]\right),$$

where

$$A_{5,\chi} = \prod_{p \neq 5}\left(1 - \frac{(\chi(p)-1)^2}{(p-1)^2}\right) \approx 1.891 + 1.559i.$$

Comparison with numerics: $q = 5$

x	$\pi(x;5,(1,1))$	$\pi(x;5,(1,2))$	$\pi(x;5,(1,3))$	$\pi(x;5,(1,4))$
10^9	$2.328 \cdot 10^6$ $2.354 \cdot 10^6$	$3.842 \cdot 10^6$ $3.774 \cdot 10^6$	$3.796 \cdot 10^6$ $3.835 \cdot 10^6$	$2.745 \cdot 10^6$ $2.750 \cdot 10^6$

Comparison with numerics: $q = 5$

x	$\pi(x;5,(1,1))$	$\pi(x;5,(1,2))$	$\pi(x;5,(1,3))$	$\pi(x;5,(1,4))$
10^9	$2.328 \cdot 10^6$	$3.842 \cdot 10^6$	$3.796 \cdot 10^6$	$2.745 \cdot 10^6$
	$2.354 \cdot 10^6$	$3.774 \cdot 10^6$	$3.835 \cdot 10^6$	$2.750 \cdot 10^6$
10^{10}	$2.142 \cdot 10^7$	$3.369 \cdot 10^7$	$3.348 \cdot 10^7$	$2.516 \cdot 10^7$
	$2.164 \cdot 10^7$	$3.324 \cdot 10^7$	$3.374 \cdot 10^7$	$2.515 \cdot 10^7$

Comparison with numerics: $q = 5$

x	$\pi(x;5,(1,1))$	$\pi(x;5,(1,2))$	$\pi(x;5,(1,3))$	$\pi(x;5,(1,4))$
10^9	$2.328 \cdot 10^6$	$3.842 \cdot 10^6$	$3.796 \cdot 10^6$	$2.745 \cdot 10^6$
	$2.354 \cdot 10^6$	$3.774 \cdot 10^6$	$3.835 \cdot 10^6$	$2.750 \cdot 10^6$
10^{10}	$2.142 \cdot 10^7$	$3.369 \cdot 10^7$	$3.348 \cdot 10^7$	$2.516 \cdot 10^7$
	$2.164 \cdot 10^7$	$3.324 \cdot 10^7$	$3.374 \cdot 10^7$	$2.515 \cdot 10^7$
10^{11}	$1.984 \cdot 10^8$	$3.000 \cdot 10^8$	$2.993 \cdot 10^8$	$2.318 \cdot 10^8$
	$2.002 \cdot 10^8$	$2.969 \cdot 10^8$	$3.011 \cdot 10^8$	$2.314 \cdot 10^8$

Comparison with numerics: $q = 5$

x	$\pi(x;5,(1,1))$	$\pi(x;5,(1,2))$	$\pi(x;5,(1,3))$	$\pi(x;5,(1,4))$
10^9	$2.328 \cdot 10^6$	$3.842 \cdot 10^6$	$3.796 \cdot 10^6$	$2.745 \cdot 10^6$
	$2.354 \cdot 10^6$	$3.774 \cdot 10^6$	$3.835 \cdot 10^6$	$2.750 \cdot 10^6$
10^{10}	$2.142 \cdot 10^7$	$3.369 \cdot 10^7$	$3.348 \cdot 10^7$	$2.516 \cdot 10^7$
	$2.164 \cdot 10^7$	$3.324 \cdot 10^7$	$3.374 \cdot 10^7$	$2.515 \cdot 10^7$
10^{11}	$1.984 \cdot 10^8$	$3.000 \cdot 10^8$	$2.993 \cdot 10^8$	$2.318 \cdot 10^8$
	$2.002 \cdot 10^8$	$2.969 \cdot 10^8$	$3.011 \cdot 10^8$	$2.314 \cdot 10^8$
10^{12}	$1.848 \cdot 10^9$	$2.704 \cdot 10^9$	$2.706 \cdot 10^9$	$2.145 \cdot 10^9$
	$1.863 \cdot 10^9$	$2.682 \cdot 10^9$	$2.717 \cdot 10^9$	$2.141 \cdot 10^9$

More on the conjectures for $r = 2$

If $a = b$ then

$$\pi(x; q, (a, a)) \sim \frac{\mathrm{li}(x)}{\phi(q)^2}\Big(1 - \frac{\phi(q) - 1}{2}\frac{\log\log x}{\log x} + \Big(\phi(q)\log\frac{q}{2\pi} + \log 2\pi - \phi(q)\sum_{p|q}\frac{\log p}{p-1}\Big)\frac{1}{2\log x}\Big).$$

More on the conjectures for $r = 2$

If $a = b$ then

$$\pi(x; q, (a, a)) \sim \frac{\mathrm{li}(x)}{\phi(q)^2}\Big(1 - \frac{\phi(q) - 1}{2}\frac{\log\log x}{\log x} + \Big(\phi(q)\log\frac{q}{2\pi} + \log 2\pi - \phi(q)\sum_{p|q}\frac{\log p}{p-1}\Big)\frac{1}{2\log x}\Big).$$

If $a \neq b$ then

$$\pi(x; q, (a, b)) \sim \frac{\mathrm{li}(x)}{\phi(q)^2}\Big(1 + \frac{1}{2}\frac{\log\log x}{\log x} + \frac{c_2(q; (a, b))}{\log x}\Big).$$

More on the conjectures for $r = 2$

If $a = b$ then

$$\pi(x;q,(a,a)) \sim \frac{\mathrm{li}(x)}{\phi(q)^2}\Big(1 - \frac{\phi(q)-1}{2}\frac{\log\log x}{\log x} + \Big(\phi(q)\log\frac{q}{2\pi} + \log 2\pi - \phi(q)\sum_{p|q}\frac{\log p}{p-1}\Big)\frac{1}{2\log x}\Big).$$

If $a \neq b$ then

$$\pi(x;q,(a,b)) \sim \frac{\mathrm{li}(x)}{\phi(q)^2}\Big(1 + \frac{1}{2}\frac{\log\log x}{\log x} + \frac{c_2(q;(a,b))}{\log x}\Big).$$

Here c_2 is complicated, but

$$c_2(q;(a,b)) + c_2(q;(b,a)) = \log(2\pi) - \phi(q)\frac{\Lambda(q/(q,b-a))}{\phi(q/(q,b-a))}.$$

Other consequences

Conjecture

Let q be prime. For large x

$$\sum_{p_n \le x} \left(\frac{p_n p_{n+1}}{q}\right) \sim -\frac{\mathrm{li}(x)}{2\log x} \log\left(\frac{2\pi \log x}{q}\right).$$

Other consequences

Conjecture

Let q be prime. For large x

$$\sum_{p_n \leq x} \left(\frac{p_n p_{n+1}}{q} \right) \sim -\frac{\mathrm{li}(x)}{2\log x} \log\left(\frac{2\pi \log x}{q} \right).$$

Conjecture

$\#\{p_n \leq x : p_n \equiv p_{n+2} (mod\ q)\}$

Other consequences

Conjecture

Let q be prime. For large x

$$\sum_{p_n \le x} \left(\frac{p_n p_{n+1}}{q}\right) \sim -\frac{\mathrm{li}(x)}{2\log x} \log\left(\frac{2\pi \log x}{q}\right).$$

Conjecture

$$\#\{p_n \le x : p_n \equiv p_{n+2} (mod\ q)\} \sim \frac{\mathrm{li}(x)}{\phi(q)}\left(1 - \frac{\phi(q)-1}{2}\frac{1}{\log x}\right).$$

The heuristic when $r = 2$

The heuristic when $r = 2$

$$\pi(x; q, (a, b)) =$$

The heuristic when $r = 2$

$$\pi(x; q, (a, b)) = \sum_{\substack{n<x: \\ n \equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n)$$

The heuristic when $r = 2$

$$\pi(x; q, (a,b)) = \sum_{\substack{n<x: \\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0: \\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h)$$

The heuristic when $r = 2$

$$\pi(x; q, (a,b)) = \sum_{\substack{n<x: \\ n\equiv a (\mathrm{mod}\ q)}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0: \\ h\equiv b-a (\mathrm{mod}\ q)}} \mathbf{1}_{\mathcal{P}}(n+h) \cdot$$

$$\cdot \prod_{\substack{t<h: \\ (t+a,q)=1}} \Big(1 - \mathbf{1}_{\mathcal{P}}(n+t)\Big)$$

The heuristic when $r = 2$

$$\pi(x; q, (a, b)) \approx \sum_{\substack{n<x:\\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0:\\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h) \cdot$$

$$\cdot \prod_{\substack{t<h:\\ (t+a,q)=1}} \left(1 - \frac{1}{\log x}\right)$$

Please do not try this at home!

The heuristic when $r = 2$

$$\pi(x; q, (a, b)) \approx \sum_{\substack{n<x:\\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0:\\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h) \cdot$$

$$\cdot \prod_{\substack{t<h:\\ (t+a,q)=1}} \left(1 - \frac{q}{\phi(q)} \frac{1}{\log x}\right)$$

The heuristic when $r = 2$

$$\pi(x;q,(a,b)) \approx \sum_{\substack{n<x:\\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0:\\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h) \cdot$$

$$\cdot \prod_{\substack{t<h:\\ (t+a,q)=1}} \left(1 - \frac{q}{\phi(q)}\frac{1}{\log x}\right)$$

$$\sim \sum_{\substack{n<x:\\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0:\\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h) e^{-h/\log x}$$

The heuristic when $r = 2$

$$\pi(x; q, (a,b)) \approx \sum_{\substack{n<x: \\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0: \\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h) \cdot$$

$$\cdot \prod_{\substack{t<h: \\ (t+a,q)=1}} \Big(1 - \frac{q}{\phi(q)} \frac{1}{\log x}\Big)$$

$$\sim \sum_{\substack{n<x: \\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \sum_{\substack{h>0: \\ h\equiv b-a \pmod q}} \mathbf{1}_{\mathcal{P}}(n+h) e^{-h/\log x}$$

$$= \sum_{h\equiv b-a \pmod q} e^{-h/\log x} \sum_{\substack{n<x: \\ n\equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n) \mathbf{1}_{\mathcal{P}}(n+h)$$

The Hardy-Littlewood conjecture

We need to understand

$$\sum_{h \equiv b-a \pmod{q}} e^{-h/\log x} \sum_{\substack{n<x \\ n \equiv a \pmod{q}}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h).$$

The Hardy-Littlewood conjecture

We need to understand

$$\sum_{h \equiv b-a \pmod q} e^{-h/\log x} \sum_{\substack{n<x \\ n \equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h).$$

Conjecture (Hardy-Littlewood)

For any $h \neq 0$, we have

$$\sum_{n<x} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h) \sim \mathfrak{S}(h)\frac{x}{\log^2 x},$$

The Hardy-Littlewood conjecture

We need to understand

$$\sum_{h \equiv b-a \pmod q} e^{-h/\log x} \sum_{\substack{n<x \\ n \equiv a \pmod q}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h).$$

Conjecture (Hardy-Littlewood)

For any $h \neq 0$, we have

$$\sum_{n<x} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h) \sim \mathfrak{S}(h)\frac{x}{\log^2 x},$$

where

$$\mathfrak{S}(h) = \prod_p \left(1 - \frac{\#(\{0,h\} \, mod \, p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-2}.$$

The Hardy-Littlewood conjecture

We need to understand

$$\sum_{h \equiv b-a (\text{mod } q)} e^{-h/\log x} \sum_{\substack{n<x \\ n \equiv a (\text{mod } q)}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h).$$

Conjecture (Hardy-Littlewood)

For any $h \neq 0$, we have

$$\sum_{\substack{n<x \\ n \equiv a (mod\ q)}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h) \sim \mathfrak{S}(h)\frac{x}{\log^2 x},$$

where

$$\mathfrak{S}(h) = \prod_p \left(1 - \frac{\#(\{0,h\} mod\, p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-2}.$$

The Hardy-Littlewood conjecture

We need to understand

$$\sum_{h \equiv b-a (\text{mod } q)} e^{-h/\log x} \sum_{\substack{n<x \\ n \equiv a (\text{mod } q)}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h).$$

Conjecture (Hardy-Littlewood)

For any $h \neq 0$ with $(h+a, q) = 1$, we have

$$\sum_{\substack{n<x \\ n \equiv a (mod\ q)}} \mathbf{1}_{\mathcal{P}}(n)\mathbf{1}_{\mathcal{P}}(n+h) \sim \frac{q}{\phi(q)^2} \mathfrak{S}_q(h) \frac{x}{\log^2 x},$$

where

$$\mathfrak{S}_q(h) = \prod_{p \nmid q} \left(1 - \frac{\#(\{0,h\} mod\, p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-2}.$$

The main term

We now have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b - a \pmod{q}} \mathfrak{S}_q(h) e^{-h/\log x}.$$

The main term

We now have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod{q}} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Gallagher: $\mathfrak{S}_q(h) = 1$ on average

The main term

We now have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod{q}} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Gallagher: $\mathfrak{S}_q(h) = 1$ on average

Blindly plugging this in, we have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod{q}} e^{-h/\log x}$$

The main term

We now have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a (\mathrm{mod}\ q)} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Gallagher: $\mathfrak{S}_q(h) = 1$ on average

Blindly plugging this in, we have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a (\mathrm{mod}\ q)} e^{-h/\log x}$$
$$\sim \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \frac{\log x}{q}$$

The main term

We now have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod q} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Gallagher: $\mathfrak{S}_q(h) = 1$ on average

Blindly plugging this in, we have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod q} e^{-h/\log x}$$

$$\sim \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \frac{\log x}{q} = \frac{1}{\phi(q)^2} \frac{x}{\log x},$$

The main term

We now have

$$\pi(x; q, (a, b)) \approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod{q}} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Gallagher: $\mathfrak{S}_q(h) = 1$ on average

Blindly plugging this in, we have

$$\begin{aligned}\pi(x; q, (a, b)) &\approx \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \sum_{h \equiv b-a \pmod{q}} e^{-h/\log x} \\ &\sim \frac{q}{\phi(q)^2} \frac{x}{\log^2 x} \frac{\log x}{q} = \frac{1}{\phi(q)^2} \frac{x}{\log x},\end{aligned}$$

which is the main term.

The source of the bias

Consider

$$\sum_{h \equiv b-a \pmod q} \mathfrak{S}_q(h) e^{-h/\log x}.$$

The source of the bias

Consider

$$\sum_{h \equiv b-a (\text{mod } q)} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Proposition

We have

$$\sum_{h \equiv 0 (mod\ q)} \mathfrak{S}_q(h) e^{-h/\log x} = \frac{\log x}{q} - \frac{\phi(q)}{2q} \log\log x + O(1),$$

The source of the bias

Consider

$$\sum_{h \equiv b-a (\mathrm{mod}\ q)} \mathfrak{S}_q(h) e^{-h/\log x}.$$

Proposition

We have

$$\sum_{h \equiv 0 (mod\ q)} \mathfrak{S}_q(h) e^{-h/\log x} = \frac{\log x}{q} - \frac{\phi(q)}{2q} \log\log x + O(1),$$

while for $v \not\equiv 0 (mod\ q)$*,*

$$\sum_{h \equiv v (mod\ q)} \mathfrak{S}_q(h) e^{-h/\log x} = \frac{\log x}{q} + O(1).$$

The source of the bias

Consider

$$\sum_{h\equiv b-a(\mathrm{mod}\ q)} \mathfrak{S}_q(h)e^{-h/\log x}.$$

Proposition

We have

$$\sum_{h\equiv 0(mod\ q)} \mathfrak{S}_q(h)e^{-h/\log x} = \frac{\log x}{q} - \frac{\phi(q)}{2q}\log\log x + O(1),$$

while for $v \not\equiv 0(mod\ q)$,

$$\sum_{h\equiv v(mod\ q)} \mathfrak{S}_q(h)e^{-h/\log x} = \frac{\log x}{q} + O(1).$$

Idea

Only the first Dirichlet series has a pole at $s = 0$.

Much needed rigor

To do this properly, we need to be more careful with

$$\prod_{\substack{t<h: \\ (t+a,q)=1}} (1 - \mathbf{1}_{\mathcal{P}}(n+t)) \approx \prod_{\substack{t<h: \\ (t+a,q)=1}} \left(1 - \frac{q}{\phi(q)\log x}\right).$$

Much needed rigor

To do this properly, we need to be more careful with

$$\prod_{\substack{t<h:\\(t+a,q)=1}} (1-\mathbf{1}_{\mathcal{P}}(n+t)) \approx \prod_{\substack{t<h:\\(t+a,q)=1}} \left(1-\frac{q}{\phi(q)\log x}\right).$$

Could expand out and invoke Hardy-Littlewood,

Much needed rigor

To do this properly, we need to be more careful with

$$\prod_{\substack{t<h:\\(t+a,q)=1}} (1 - \mathbf{1}_{\mathcal{P}}(n+t)) \approx \prod_{\substack{t<h:\\(t+a,q)=1}} \left(1 - \frac{q}{\phi(q)\log x}\right).$$

Could expand out and invoke Hardy-Littlewood, but this is ugly!

Much needed rigor

To do this properly, we need to be more careful with

$$\prod_{\substack{t<h:\\(t+a,q)=1}} (1 - \mathbf{1}_{\mathcal{P}}(n+t)) \approx \prod_{\substack{t<h:\\(t+a,q)=1}} \left(1 - \frac{q}{\phi(q)\log x}\right).$$

Could expand out and invoke Hardy-Littlewood, but this is ugly!

Better idea: Incorporate inclusion-exclusion directly into H-L.

Modified Hardy-Littlewood

Define $\widetilde{\mathbf{1}}_{\mathcal{P}}(n) = \mathbf{1}_{\mathcal{P}}(n) - \frac{q}{\phi(q)\log n}$.

Modified Hardy-Littlewood

Define $\widetilde{\mathbf{1}}_{\mathcal{P}}(n) = \mathbf{1}_{\mathcal{P}}(n) - \frac{q}{\phi(q)\log n}$.

Conjecture

If $|\mathcal{H}| = k$ *with* $(h + a, q) = 1$ *for all* $h \in \mathcal{H}$*, then*

$$\sum_{\substack{n \le x \\ n \equiv a (mod\ q)}} \prod_{h \in \mathcal{H}} \widetilde{\mathbf{1}}_{\mathcal{P}}(n + h) \sim \frac{q^{k-1}}{\phi(q)^k} \mathfrak{S}_{q,0}(\mathcal{H}) \frac{x}{\log^k x},$$

Modified Hardy-Littlewood

Define $\widetilde{\mathbf{1}}_{\mathcal{P}}(n) = \mathbf{1}_{\mathcal{P}}(n) - \frac{q}{\phi(q)\log n}$.

Conjecture

If $|\mathcal{H}| = k$ *with* $(h+a, q) = 1$ *for all* $h \in \mathcal{H}$, *then*

$$\sum_{\substack{n \le x \\ n \equiv a (mod\ q)}} \prod_{h \in \mathcal{H}} \widetilde{\mathbf{1}}_{\mathcal{P}}(n+h) \sim \frac{q^{k-1}}{\phi(q)^k} \mathfrak{S}_{q,0}(\mathcal{H}) \frac{x}{\log^k x},$$

where

$$\mathfrak{S}_{q,0}(\mathcal{H}) := \sum_{\mathcal{T} \subseteq \mathcal{H}} (-1)^{|\mathcal{H} \setminus \mathcal{T}|} \mathfrak{S}_q(\mathcal{T}).$$

Sums of modified singular series

Theorem (Montgomery, S)

$$\sum_{\substack{\mathcal{H}\subseteq[1,h]\\|\mathcal{H}|=k}} \mathfrak{S}_0(\mathcal{H}) = \frac{\mu_k}{k!}(-h\log h + Ah)^{k/2} + O_k(h^{k/2-\delta}),$$

where $\mu_k = 0$ *if* k *is odd.*

Sums of modified singular series

Theorem (Montgomery, S)

$$\sum_{\substack{\mathcal{H}\subseteq[1,h]\\|\mathcal{H}|=k}} \mathfrak{S}_0(\mathcal{H}) = \frac{\mu_k}{k!}(-h\log h + Ah)^{k/2} + O_k(h^{k/2-\delta}),$$

where $\mu_k = 0$ if k is odd.

Point

We can discard $\mathcal{H}$ with $|\mathcal{H}| \geq 3$.

Notes on assembly

Three flavors of two-term $\mathcal{H}$:

Notes on assembly

Three flavors of two-term $\mathcal{H}$:

- $\mathcal{H} = \{0, h\}$: Contributes main-term, bias against diagonal

Notes on assembly

Three flavors of two-term $\mathcal{H}$:

- $\mathcal{H} = \{0, h\}$: Contributes main-term, bias against diagonal
- $\mathcal{H} = \{0, t\}$, $\{t, h\}$: (One interloper) Contributes to c_2

Notes on assembly

Three flavors of two-term $\mathcal{H}$:

- $\mathcal{H} = \{0, h\}$: Contributes main-term, bias against diagonal
- $\mathcal{H} = \{0, t\}$, $\{t, h\}$: (One interloper) Contributes to c_2
- $\mathcal{H} = \{t_1, t_2\}$: (Two interlopers) Unbiased

Notes on assembly

Three flavors of two-term $\mathcal{H}$:

- $\mathcal{H} = \{0, h\}$: Contributes main-term, bias against diagonal
- $\mathcal{H} = \{0, t\}$, $\{t, h\}$: (One interloper) Contributes to c_2
- $\mathcal{H} = \{t_1, t_2\}$: (Two interlopers) Unbiased

Remark

We expect $\mathcal{H}$ with $|\mathcal{H}| \geq 3$ to contribute further lower-order terms.

The conjecture

Conjecture (LO & S)

Let $\mathbf{a} = (a_1, \ldots, a_r)$ *with* $r \geq 2$. *Then*

$$\pi(x; q, \mathbf{a}) = \frac{\mathrm{li}(x)}{\phi(q)^r} \left[1 + c_1(q; \mathbf{a}) \frac{\log \log x}{\log x} + \frac{c_2(q; \mathbf{a})}{\log x} + O\left(\log^{-7/4} x \right) \right],$$

where

$$c_1(q; \mathbf{a}) = \frac{\phi(q)}{2} \left(\frac{r-1}{\phi(q)} - \#\{1 \leq i < r : a_i \equiv a_{i+1} (mod\ q)\} \right),$$

and $c_2(q; \mathbf{a})$ *is complicated but explicit.*