BIG BIASES AMONGST PRODUCTS OF TWO PRIMES

DAVID DUMMIT, ANDREW GRANVILLE, AND HERSHY KISILEVSKY

ABSTRACT. We show that substantially more than a quarter of the odd integers of the form pq up to x, with p, q both prime, satisfy $p \equiv q \equiv 3 \pmod 4$.

1. Introduction

There are roughly equal quantities of odd integers, that are the product of two primes, in the arithmetic progressions 1 (mod 4) and 3 (mod 4). Indeed the counts differ by no more than $x^{1/2+o(1)}$ (assuming the Riemann Hypothesis for $L(1, (-4/.))$; see [1] for a detailed analysis). One might guess that these integers $pq \le x$ are further evenly split amongst those with p and q in pre-specified arithmetic progressions mod 4, but recent calculations reveal a substantial bias towards those $pq \le x$ with $p \equiv q \equiv 3 \pmod 4$. Indeed for the ratio

$$r(x) := \#\{pq \le x : \ p \equiv q \equiv 3 \pmod 4\} \Big/ \frac{1}{4}\#\{pq \le x\}$$

we found that

$$r(1000) \approx 1.347, \ r(10^4) \approx 1.258, \ r(10^5) \approx 1.212, \ r(10^6) \approx 1.183, \ r(10^7) \approx 1.162,$$

a pronounced bias that seems to be converging to 1 surprisingly slowly. We will show that this is no accident and that there is similarly slow convergence for many such questions:

Theorem 1.1. *Let χ be a quadratic character of conductor d. For $\eta = -1$ or 1 we have*

$$\frac{\#\{pq \le x : \ \chi(p) = \chi(q) = \eta\}}{\frac{1}{4}\#\{pq \le x : \ (pq, d) = 1\}} = 1 + \eta\frac{(\mathcal{L}_\chi + o(1))}{\log\log x} \quad \textit{where} \quad \mathcal{L}_\chi := \sum_p \frac{\chi(p)}{p}.$$

If $\chi = (-4/.)$ then $\mathcal{L}_\chi = -.334\ldots$ so the theorem implies that $r(x) \ge 1 + \frac{(1+o(1))}{3(\log\log x - 1)}$. If we let $s(x) = 1 + \frac{1}{3(\log\log x - 1)}$ then we have

$$s(1000) \approx 1.357, \ s(10^4) \approx 1.273, \ s(10^5) \approx 1.230, \ s(10^6) \approx 1.205, \ s(10^7) \approx 1.187,$$

a pretty good fit with the data above. The prime numbers have only been computed up to something like 10^{24} so it is barely feasible that one could collect data on this problem up to 10^{50} in the foreseeable future. Therefore we would expect this bias to be at least 7% on any data that will be collected this century (as $s(10^{50}) \approx 1.07$).

Date: March 25, 2015.
2010 *Mathematics Subject Classification.* 11L20, 11N25, 11N13.
Key words and phrases. Primes in arithmetic progressions, Prime races.
We would like to thank the anonymous referee for some helpful corrections.

Proof. For a given quadratic Dirichlet character χ we will count the number of integers $pq \le x$ with $\chi(p) = \chi(q) = 1$ (and the analogous argument works for -1). One can write any such integer $pq \le x$ with $p \le q \le x/p$, so that $p \le \sqrt{x}$. Hence we wish to determine

$$\sum_{\substack{p \le \sqrt{x} \\ \chi(p)=1}} \sum_{\substack{p \le q \le x/p \\ \chi(q)=1}} 1. \tag{1}$$

We will use the prime number theorem for arithmetic progressions in the form

$$\sum_{\substack{q \le Q \\ \chi(q)=1 \\ q \text{ prime}}} 1 = \frac{Q}{2\log Q} + O\left(\frac{Q}{(\log Q)^2}\right), \tag{2}$$

as well as the same estimate for the number of primes $q \le Q$ with $\chi(q) = -1$. Therefore the sum in (1) equals

$$\sum_{p \le \sqrt{x}} \left\{ \frac{(\chi_0(p) + \chi(p))}{2} \cdot \frac{x}{2p\log(x/p)} + O\left(\frac{x}{p(\log x)^2} + \frac{p}{\log p}\right) \right\}$$

where the implicit constant in the $O(.)$ depends only on the conductor d of χ, and χ_0 is the principal character $\pmod d$. This equals

$$\frac{1}{4} \sum_{\substack{p \le \sqrt{x} \\ (p,d)=1}} \frac{x}{p\log(x/p)} + \frac{x}{4} \sum_{p \le \sqrt{x}} \frac{\chi(p)}{p\log(x/p)} + O\left(\frac{x}{(\log x)^2} \log\log x\right).$$

The difference between the second sum, and the same sum with $\log(x/p)$ replaced by $\log x$, is

$$\frac{x}{4\log x} \sum_{p \le \sqrt{x}} \frac{\chi(p)\log p}{p\log(x/p)} \ll \frac{x}{(\log x)^2} \log\log x.$$

using the prime number theorem for arithmetic progressions (as in (2)) and partial summation. These concepts also imply that

$$\frac{x}{4\log x} \sum_{p > \sqrt{x}} \frac{\chi(p)}{p} \ll \frac{x}{(\log x)^2}.$$

Collecting together what we have proved so far yields that $\#\{pq \le x : \ \chi(p) = \chi(q) = 1\}$

$$= \frac{1}{4} \left\{ \#\{pq \le x : \ (p,d) = 1\} + \frac{x}{\log x} \sum_p \frac{\chi(p)}{p} + O\left(\frac{x}{(\log x)^2} \log\log x\right) \right\}$$

The first term is well-known to equal $\frac{x}{\log x}(\log\log x + O(1))$, and so we deduce that

$$\frac{\#\{pq \le x : \ \chi(p) = \chi(q) = 1\}}{\frac{1}{4}\#\{pq \le x : \ (pq,d) = 1\}} = 1 + \frac{1}{\log\log x} \left(\sum_p \frac{\chi(p)}{p} + o(1) \right).$$

as claimed. ∎

We note that

$$\sum_p \frac{\chi(p)}{p} = \sum_{m\geq 1} \frac{\mu(m)}{m} \log L(m, \chi^m) = \log L(1,\chi) + E(\chi),$$

where

$$\sum_p \left(\log\left(1 - \frac{1}{p}\right) + \frac{1}{p}\right) = -0.315718\ldots \leq E(\chi) \leq \sum_p \left(\log\left(1 + \frac{1}{p}\right) - \frac{1}{p}\right) = -0.18198\ldots$$

2. Further remarks

• One deduces from our theorem that $r(x) > 1$ for all x sufficiently large. We conjecture that this is true for all $x \geq 9$.

• We also conjecture that $\mathcal{L}_\chi$ is non-zero for all quadratic characters χ, so that our Theorem implies that there is always such a bias. We would further conjecture that $\mathcal{L}_\chi$ is non-zero for all non-principal characters χ.

• One can calculate the bias in other such questions. For example, we get roughly triple the bias for the proportion of $pq \leq x$ for which $\left(\frac{p}{5}\right) = \left(\frac{q}{5}\right) = -1$ out of all $pq \leq x$ with $p, q \neq 5$, since $\mathcal{L}_{(./5)} \approx -1.008$. The data

$$r_5(1000) \approx 1.881,\ r_5(10^4) \approx 1.626,\ r_5(10^5) \approx 1.523,\ r_5(10^6) \approx 1.457,\ r_5(10^7) \approx 1.416,$$

confirms this very substantial bias. It would be interesting to find more extreme examples.

• How large can the bias get if $d \leq x$? It is known [2] that $L(1, \chi)$ can be as large as $c \log\log d$, and so $\mathcal{L}_\chi$ can be as large as $\log\log\log d + O(1)$. We conjecture that there exists $d \leq x$ for which the bias in our Theorem is as large as

$$1 + \frac{\log\log\log x + O(1)}{\log\log x}.$$

Note that this requires proving a uniform version of the Theorem. Our proof here is not easily modified to resolve this problem, since it assumes that x is taken to be very large compared to d.

• The same bias can be seen (for much the same reason) in looking at

$$\sum_{\substack{p\leq x\\ p\equiv 3 \pmod 4}} \frac{1}{p} \Bigg/ \sum_{\substack{p\leq x\\ p\equiv 1 \pmod 4}} \frac{1}{p} \approx 1 + \frac{2}{3\log\log x}.$$

Indeed, by the analogous proof, we have in general

$$\sum_{\substack{p\leq x\\ \chi(p)=1}} \frac{1}{p} \Bigg/ \sum_{\substack{p\leq x\\ \chi(p)=-1}} \frac{1}{p} = 1 + 2\frac{(\mathcal{L}_\chi + o(1))}{\log\log x}.$$

We therefore see a bias in the distribution of primes in arithmetic progressions, where each prime p is weighted by $1/p$, corresponding to the sign of $\mathcal{L}_\chi$. This effect is much more pronounced than in the traditional prime race problem where the same comparison is made, though with each prime weighted by 1. The bias here is determined by the distribution of values of $\chi(p)$, whereas the prime race bias is determined by the values of $\chi(p^2) = 1$, so they appear to be independent phenomena. However one might guess that both biases are sensitive to low lying zeros of $L(s,\chi)$. This probably deserves further investigation, to determine whether there are any correlations between the two biases.

3. Generalizations

The proof of the theorem generalizes to show that, for given quadratic characters $\chi_1,\dots,\chi_k$, and $(\eta_1,\eta_2\dots,\eta_k)\in\{-1,1\}^k$, the proportion of the k-tuples of primes $p_1<p_2<\dots<p_k$ with $p_1\cdots p_k\le x$, which satisfy $\chi_j(p_j)=\eta_j$ for each $j,\ 1\le j\le k$, equals

$$2^{-k}\left(1+\frac{\eta_1\mathcal{L}_{\chi_1}+o(1)}{\log\log x}\right)\quad\text{as}\quad x\to\infty.$$

Allowing any ordering of the prime factors p_j, we deduce that the proportion of the k-tuples of primes $p_1,p_2,\dots,p_k$ with $p_1\cdots p_k\le x$, which satisfy $\chi_j(p_j)=\eta_j$ for each j, equals

$$2^{-k}\left(1+\frac{c(\vec{\chi},\vec{\eta})+o(1)}{\log\log x}\right)\quad\text{as}\quad x\to\infty,\quad\text{where}\quad c(\vec{\chi},\vec{\eta}):=\frac{1}{k}\sum_{j=1}^{k}\eta_j\mathcal{L}_{\chi_j}.$$

This tends to the expected proportion 2^{-k} as $x\to\infty$, but exhibits that there is a substantial bias up to any point up to which one might feasibly calculate, provided $c(\vec{\chi},\vec{\eta})\ne 0$.

• There is no such bias (i.e. $c(\vec{\chi},\vec{\eta})=0$) when $k=2,\ \chi_1=\chi_2$ and $\eta_1+\eta_2=0$. Can one prove that $c(\vec{\chi},\vec{\eta})$ can only be 0 for such trivial reasons? That is, is $c(\vec{\chi},\vec{\eta})=0$ if and only if $\sum_{j:\ \chi_j=\chi}\eta_j=0$ for every character $\chi\in\vec{\chi}$?

• We deduce, from the last displayed equation, that the proportion of the integers $n\le x$ with exactly k distinct prime factors, which satisfy $\chi(p)=\eta$ for each prime p dividing n, equals

$$2^{-k}\left(1+\frac{\eta\mathcal{L}_{\chi}+o(1)}{\log\log x}\right)\quad\text{as}\quad x\to\infty.$$

That is, we have the same bias, no matter how many prime factors n has. This proof works for k fixed as $x\to\infty$. It would be interesting to understand the bias if k gets large with x, particularly when $k\sim\log\log x$, the typical number of prime factors of an integer $\le x$.

• Given arithmetic progressions $a \pmod m$ and $b \pmod n$, one can surely prove that there exists $\beta=\beta(a \pmod m, b \pmod n)$ such that

$$\frac{\#\{pq\le x:\ p\equiv a \pmod m,\ q\equiv b \pmod n\}}{\frac{1}{\phi(m)\phi(n)}\#\{pq\le x:\ (p,m)=(q,n)=1\}}=1+\frac{\beta+o(1)}{\log\log x}.$$

It would be interesting to classify when $\beta(a \pmod m, b \pmod n)$ is non-zero, and to determine situations in which it is large.

• More generally for non-empty subsets $A\subseteq(\mathbb{Z}/m\mathbb{Z})^*$ and $B\subseteq(\mathbb{Z}/n\mathbb{Z})^*$, there presumably exists a constant $\beta=\beta(A,B)$ for which

$$\frac{\#\{pq\le x:\ p \pmod m\in A,\ q \pmod n\in B\}}{\frac{|A|}{\phi(m)}\frac{|B|}{\phi(n)}\#\{pq\le x:\ (p,m)=(q,n)=1\}}=1+\frac{\beta+o(1)}{\log\log x}.$$

We would guess that there is no bias, that is $\beta(A,B)=0$, only if either

(i) A and B both contain all congruence classes (that is, every prime not dividing mn can be represented by both A and B); or

(ii) $A \cup B$ is a partition of the integers coprime to mn (that is, every prime not dividing mn is represented by A, or represented by B, but not both).

References

[1] Kevin Ford and Jason Sneed, *Chebyshev's bias for products of two primes*, Experiment. Math. **19** (2010), 385-398.

[2] Andrew Granville and K. Soundararajan, *The distribution of values of* $L(1, \chi_d)$, Geom. Funct. Anal. **13** (2003), 992-1028.

Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05401, USA

E-mail address: dummit@math.uvm.edu

Département de mathématiques et de statistiques,, Université de Montréal, CP 6128 succ. Centre-Ville, Montréal, QC H3C 3J7, Canada

E-mail address: andrew@dms.umontreal.ca

Department of Mathematics and Statistics, Sir George Williams Campus, Concordia University, Montreal, QC H3G 1M8, Canada

E-mail address: kisilev@mathstat.concordia.ca