Max-Planck-Institut für Mathematik
Bonn

Products of two proportional primes

by

Pieter Moree
Sumaia Saad Eddin

Max-Planck-Institut für Mathematik
Preprint Series 2016 (29)

Products of two proportional primes

Pieter Moree
Sumaia Saad Eddin

Max-Planck-Institut für Mathematik
Vivatsgasse 7
53111 Bonn
Germany

Institute of Financial Mathematics and
Applied Number Theory
Altenbergerstrasse 69
4040 Linz
Austria

MPIM 16-29

PRODUCTS OF TWO PROPORTIONAL PRIMES

PIETER MOREE AND SUMAIA SAAD EDDIN

Abstract. In RSA cryptography numbers of the form pq, with p and q two distinct proportional primes play an important role. For a fixed real number $r > 1$ we formalize this by saying that an integer pq is an RSA-integer if p and q are primes satisfying $p < q \le rp$. Recently Dummit, Granville and Kisilevsky showed that substantially more than a quarter of the odd integers of the form pq up to x, with p, q both prime, satisfy $p \equiv q \equiv 3 \pmod 4$. In this paper we investigate this phenomenon for RSA-integers. We establish an analogue of a strong form of the prime number theorem with the logarithmic integral replaced by a variant. From this we derive an asymptotic formula for the number of RSA-integers $\le x$ which is much more precise than an earlier one derived by Decker and Moree in 2008.

1. Introduction

Let $\omega(n)$ and $\Omega(n)$ denote the number of distinct, respectively total number of prime factors of n. Put

$$\pi(x,k) = \sum_{\substack{n\le x\\ \omega(n)=k}} 1 \quad \text{and} \quad N(x,k) = \sum_{\substack{n\le x\\ \Omega(n)=k}} 1.$$

The following asymptotic formula is due to Landau [9, p. 211]:

$$\pi(x,k) \sim N(x,k) \sim \frac{x}{\log x}\cdot\frac{(\log\log x)^{k-1}}{(k-1)!}. \tag{1}$$

For a nice survey on $\pi(x,k)$ and $N(x,k)$ up to 1987 see Hildebrand [7]. A recent contribution to the study of $\pi(x,k)$ is the discovery of **bias**. Define

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

(Here and in the sequel the notation p and q is exclusively used to indicate primes.) Numerically it seems that consistently $r(x) > 1$. We have, e.g., $r(10^6) \approx 1.183$ and $r(10^7) \approx 1.162$. Dummit et al. [2] showed that

$$r(x) = 1 + \frac{(\beta + o(1))}{\log\log x}, \tag{2}$$

with $\beta \approx 0.334$. This turns out to be in pretty good agreement with the observed values. Since $\beta/\log\log x$ tends to zero so slowly, (2) shows that substantially more than a quarter of the integers $pq \le x$ satisfy $p \equiv 3 (\mathrm{mod}\ 4)$ and $q \equiv 3 (\mathrm{mod}\ 4)$.

Date: June 24, 2016.
Mathematics Subject Classification (2000). 11N37, 11Y60

1.1. **RSA-integers.** In the RSA cryptosystem, see [5, Chapter 3], integers of the form $n = p{\cdot}q$ are the main actors. The security of this system is based on the current difficulty of factoring such integers (sometimes called quasiprimes) in a reasonable time. As soon as a working quantum computer is developed, it will be the end of the RSA cryptosystem [12]. The RSA cryptosystem is known to be more easily breakable under certain special restrictions on p and q. E.g., if $|p-q|$ is small or if one of p and q is much smaller than the other ("unbalanced RSA"). In RSA practice p and q are taken to be proportional, i.e. $p < q < rp$ for some $r > 1$. This does not exclude $q - p$ from being small, but for our counting purposes this suffices.

1.2. **Bias of RSA-integers.** We study two problems in this paper. One is to determine to what extent RSA-integers are biased. If they are, we would glean a very small amount of information about their prime factorisation (provided they are generated randomly), and so the question is somewhat relevant. The other problem is to find a precise asymptotic for the counting function of RSA-integers

$$C_r(x) := \#\left\{pq \le x : \ p < q \le rp\right\},$$

where $r > 1$ is an arbitrary real fixed number.

Note that $C_r(x)$ is the RSA-analogue of $\pi(x, 2)$. Theorem 1 gives the asymptotic behaviour of $C_r(x)$. Comparison with (1) shows that there are much fewer RSA-integers than integers having two (distinct) prime factors.

Theorem 1 (Decker and Moree [1])**.** *Let $r > 1$ be a real number. As x tends to infinity we have*

$$C_r(x) = \frac{2x\log r}{\log^2 x} + \mathcal{O}\left(\frac{rx\log(er)}{\log^3 x}\right).$$

This result was generalized by Hashimoto [6] who determined the asymptotic behaviour of $\#\left\{pq \le x : \ p < q \le f(p)\right\}$ for a large class of functions f satisfying $f(x) > x$. Another generalization was obtain by Justus [8] who obtained an asymptotic for $\#\left\{pq \le x : \ p < q \le x^{\theta}p\right\}$, with $0 < \theta < 1$ fixed. On the more cryptographic side, there is the dissertation by Loebenberger [10].

The main aim of this paper is to establish a very precise asymptotic formula for $C_r(x)$ (Corollary 4). On our way towards establishing this, we show that RSA-integers are rather unbiased (Corollary 2). As a particular case we obtain that the RSA-integer analogue of $r(x)$ shows little bias (Corollary 3).

As usual by $\pi(x)$ we denote the number of primes $p \le x$. We will use the prime number theorem in the form

$$\pi(x) = \mathrm{Li}(x) + \mathcal{O}\Big(xe^{-c\sqrt{\log x}}\Big), \tag{3}$$

where

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}$$

denotes the logarithmic integral.

In our main result, a variant, $F_r(x)$, of the logarithmic integral will play the main role. It is easily seen to be a **concave** function for $x \ge 2r$.

Theorem 2. *Let $r > 1$ be an arbitrary fixed real number. Given two sets of primes S_1 and S_2, we put*

$$D_r(x) := \#\{pq \le x : p < q \le rp, \ p \in S_1, \ q \in S_2\}.$$

Suppose that for $j = 1, 2$ the counting functions associated to S_j satisfy

$$\pi_{S_j}(x) := \sum_{\substack{p \le x \\ p \in S_j}} 1 = \frac{1}{\delta_j}\mathrm{Li}(x) + \mathcal{O}\left(xe^{-c\sqrt{\log x}}\right), \tag{4}$$

where $\delta_j > 0$ and $c > 0$ is a positive constant. For $x \ge 2r$ put

$$F_r(x) = \int_{2r}^{x} \frac{\log\log\sqrt{rt} - \log\log\sqrt{t/r}}{\log t}dt.$$

For $x \ge 2r$ and x tending to infinity we have

$$\delta_1\delta_2 D_r(x) = F_r(x) + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}}\right),$$

where $c(\epsilon) = (1-\epsilon)c/\sqrt{2}$ and $0 < \epsilon < 1$ is arbitrary.

Corollary 1. *We have $C_r(x) = F_r(x) + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}}\right)$.*

Proof. For S_1 and S_2 we take the set of all primes. It follows by (3) that condition (4) is satisfied with $\delta_1 = \delta_2 = 1$. □

Corollary 2. *We have $\delta_1\delta_2 D_r(x) = C_r(x) + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}}\right)$ and*

$$R_r(x) := \frac{\delta_1\delta_2 D_r(x)}{C_r(x)} = 1 + \mathcal{O}_r\left((\log^2 x)\ e^{-c(\epsilon)\sqrt{\log x}}\right).$$

Proof. Follows from Theorem 2, Corollary 1 and Theorem 1. □

Corollary 3. *Let a_1, d_1, a_2, d_2 be natural numbers with $(a_1, d_1) = (a_2, d_2) = 1$. We have*

$$\frac{\#\{pq \le x : p \equiv a_1 (\mathrm{mod}\ d_1),\ q \equiv a_2 (\mathrm{mod}\ d_2),\ p < q \le rp\}}{\#\{pq \le x : p < q \le rp\}/(\varphi(d_1)\varphi(d_2))} = 1 + \mathcal{O}_r\left((\log^2 x)\ e^{-c(\epsilon)\sqrt{\log x}}\right).$$

Proof. For S_j we take in Corollary 2 the set of all primes $\equiv a_j (\mathrm{mod}\ d_j)$. The prime number theorem for arithmetic progressions in the form

$$\pi_{S_j}(x) = \frac{\mathrm{Li}(x)}{\varphi(d_j)} + \mathcal{O}\left(xe^{-c\sqrt{\log x}}\right) \tag{5}$$

then shows that condition (4) is satisfied with $\delta_j = \varphi(d_j)$ (as usual $\varphi(d)$ denotes Euler's totient function). □

On comparing (2) with Corollary 2 (or with Corollary 3 for that matter) we see that for RSA-integers there is far less bias than for integers $n \le x$ having two distinct prime factors.

The implicit error terms of results involving the sets S_j might depend on them. For notational convenience this possible dependence is not explicitly indicated.

1.3. **Asymptotic formulas for $F_r(x)$ and $D_r(x)$.** By splitting the integration range in say 2 to $\sqrt{x}$ and $\sqrt{x}$ to x, one sees that

$$\int_2^x \frac{dt}{\log^k t} = \mathcal{O}_k\left(\frac{x}{\log^k x}\right). \tag{6}$$

Using this and partial integration we infer that for every $n \ge 2$ we have

$$\mathrm{Li}(x) = \sum_{k=1}^{n-1}(k-1)!\frac{x}{\log^k x} + \mathcal{O}_n\Big(\frac{x}{\log^n x}\Big). \tag{7}$$

Theorem 3 provides the analogue of the asymptotic formula (7) for $F_r(x)$.

Theorem 3. *Let $r > 1$ be an arbitrary fixed real number and $n \geq 2$ an integer. Then*

$$F_r(x) = \sum_{k=1}^{n-1} a_k(r) \frac{x}{\log^{k+1} x} + \mathcal{O}_n \left(\frac{x \log^{2\lfloor n/2 \rfloor +1}(2r) \log r}{\log^{n+1} x} \right),$$

where

$$a_k(r) = \sum_{j=1}^{[\frac{k+1}{2}]} \frac{k!}{(2j-1)!} \frac{2 \log^{2j-1} r}{2j-1},$$

with $[x]$ the integral part of x.

k	$a_k(r)$
1	2ρ
2	4ρ
3	$12\rho + 2\rho^3/3$
4	$48\rho + 8\rho^3/3$
5	$240\rho + 40\rho^3/3 + 2\rho^5/5$
6	$1440\rho + 80\rho^3 + 12\rho^5/5$
7	$10080\rho + 560\rho^3 + 84\rho^5/5 + 2\rho^7/7$
8	$80640\rho + 4480\rho^3 + 672\rho^5/5 + 16\rho^7/7$
9	$725760\rho + 40320\rho^3 + 6048\rho^5/5 + 144\rho^7/7 + 2\rho^9/9$
10	$7257600\rho + 403200\rho^3 + 12096\rho^5 + 1440\rho^7/7 + 20\rho^9/9$

TABLE 1. The polynomial $a_k(r)$ for $k \in \{1, 2, \ldots, 10\}$ with $\rho = \log r$.

Theorem 3 when combined with Theorem 2 now yields Theorem 4.

Theorem 4. *Let S_1 and S_2 be sets of primes satisfying the condition (4). Let $r > 1$ be an arbitrary fixed real number and $n \geq 2$ be an arbitrary integer. As x tends to infinity, we have*

$$\delta_1 \delta_2 D_r(x) = \sum_{k=1}^{n-1} a_k(r) \frac{x}{\log^{k+1} x} + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}} \right) + \mathcal{O}_n \left(\frac{x \log^{2\lfloor n/2 \rfloor +1}(2r) \log r}{\log^{n+1} x} \right),$$

where $c(\epsilon)$ and $a_k(r)$ are defined in Theorem 2, respectively Theorem 3.

Corollary 4. *Let $r > 1$ be an arbitrary fixed real number and $n \geq 2$ be an arbitrary integer. As x tends to infinity, we have*

$$C_r(x) = \sum_{k=1}^{n-1} a_k(r) \frac{x}{\log^{k+1} x} + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}} \right) + \mathcal{O}_n \left(\frac{x \log^{2\lfloor n/2 \rfloor +1}(2r) \log r}{\log^{n+1} x} \right),$$

Corollary 5. *Let $B > 0$ be an arbitrary real number. Uniformly for $1 < r \leq \log^B x$ we have*

$$C_r(x) = \sum_{k=1}^{n-1} a_k(r) \frac{x}{\log^{k+1} x} + \mathcal{O}_{n,B} \left(\frac{x}{\log^{n+1} x} (\log \log x)^{2\lfloor n/2 \rfloor +2} \right),$$

Note that Corollary 4 with $n = 2$ slightly improves on Theorem 1. With more work it is possible to improve the error terms in our results in the r aspect. As this seems to be mathematically not very important, but requires considerable effort and is not beneficial for the brevity and clarity of our presentation, we have abstained from pursuing this. Indeed, if we would have ignored the r dependence altogether, our proofs would have been simpler and shorter. We point out that we want to have estimates valid for $r > 1$, not just for r large. E.g., it is true that $1+\log r = \mathcal{O}(\log r)$ as r tends to infinity, but not if we take $r > 1$. Correct in this case is $1 + \log(r) = \log(er) = \mathcal{O}(\log(2r))$.

Our proof of Theorem 4 has Theorem 2 as a starting point. We provide some more details of the proofs in Section 2 followed by the full proofs in the remaining sections.

2. Sketch of the proofs

To understand the proofs it is helpful to first get an idea of the proof of Theorem 1.

For any prime p we define $f_p(x)$ to be the number of primes q such that $pq \le x$ and $p < q \le rp$. An easy computation then yields

$$C_r(x) = \sum_{p\le x} f_p(x) = -\sum_{p\le\sqrt{x}} \pi(p) + \sum_{p\le\sqrt{x/r}} \pi(rp) + \sum_{\sqrt{x/r}<p\le\sqrt{x}} \pi\left(\frac{x}{p}\right). \tag{8}$$

The asymptotic behaviour of each of these three sums is then determined. As input not more than the prime number theorem with error $\mathcal{O}(x\log^{-3} x)$ is used (that is the estimate (7) with $n = 3$).

Our proof of Theorem 2 starts by noting that (cf. the proof of [1, Lemma 2])

$$D_r(x) = -\sum_{\substack{p\le\sqrt{x}\\ p\in S_1}} \pi_{S_2}(p) + \sum_{\substack{p\le\sqrt{x/r}\\ p\in S_1}} \pi_{S_2}(rp) + \sum_{\substack{\sqrt{x/r}<p\le\sqrt{x}\\ p\in S_1}} \pi_{S_2}\left(\frac{x}{p}\right). \tag{9}$$

The first two sums in (9) can be dealt with the same way since the second sum with $r = 1$ is the negative of the first sum. The idea is now to replace every $\pi_{S_2}(z)$ in (9) by a $\mathrm{Li}(z)/\delta_2$, thus producing a small error (by the assumption (4)) and then to interchange the order of integration and summation. In doing so terms of the form $\pi_{S_1}(z)$ appear and those we replace by $\mathrm{Li}(z)/\delta_1$ (by assumption (4) again at the cost of introducing a small error). We thus obtain an approximation for $D_r(x)$ with main term $G_r(x)/(\delta_1\delta_2)$, where

$$G_r(x) = \frac{1}{2}\mathrm{Li}(\sqrt{x})^2 - \int_{2r}^{\sqrt{rx}} \frac{\mathrm{Li}(t/r)}{\log t}dt + \int_{\sqrt{x}}^{\sqrt{rx}} \frac{\mathrm{Li}(x/t)}{\log t}dt. \tag{10}$$

Taking the derivative of $G_r(x)$ with respect to x then shows that $G_r(x) = F_r(x) + \mathcal{O}(r)$. This then completes the proof.

In Theorem 3 we try to obtain an expansion of the form $\sum_{k=1}^{n-1} g_k(r)x\log^{-k-1} x$ for $G_r(x)$, with $g_k(r)$ to be determined. The key observation now is that $G_r'(x) = F_r'(x)$ is such a simple function that an expansion of the form $\sum_{k=1}^{n-1} h_k(r)x\log^{-k-1} x$ for $G_r'(x)$ is easily found, where the $h_k(r)$ are readily determined. Subsequently one integrates this expansion termwise. This then shows that $g_k(r) = a_k(r)$ with $a_k(r)$ as defined in Theorem 3. One has to take some care to show that this termwise integration is actually allowed.

Our first approach for establishing Theorem 3, was to substitute the expansion (7) for $\mathrm{Li}(z)$ in (10), leading to the conclusion that an expansion as in Theorem 3 exists. However, in this way complicated expressions for the polynomials $a_k(r)$ are obtained. On computing various examples of those using Mathematica and studying the j-th coefficient of $a_k(r)$ as a sequence using the On-Line Encyclopedia of Integer Sequences (OEIS), we made an explicit conjecture for the coefficients of $a_k(r)$ and eventually proved it by quite a different route.

3. Proof of Theorem 2

3.1. **Some lemmas.** In the analysis of the error term of our result, we make use of the following easy estimates. The ones in part a) arise on replacing terms of the form $\pi_{S_2}(z)$ by $\mathrm{Li}(z)/\delta_2$, the ones in part b) on replacing terms of the form $\pi_{S_1}(z)$ by $\mathrm{Li}(z)/\delta_1$.

Lemma 1. *Let $c>0$, $r>1$ and $0<\epsilon<1$. Put $c(\epsilon)=(1-\epsilon)c/\sqrt{2}$.*
a) *We have*

$$\sum_{p\le\sqrt{x/r}} pe^{-c\sqrt{\log(rp)}} \le \sum_{p\le\sqrt{x}} pe^{-c\sqrt{\log p}} = \mathcal{O}\left(xe^{-c(\epsilon)\sqrt{\log x}}\right) \tag{11}$$

and

$$\sum_{\sqrt{x/r}<p\le\sqrt{x}} \frac{x}{p}e^{-c\sqrt{\log(x/p)}} = \mathcal{O}\left(xe^{-c(\epsilon)\sqrt{\log x}}\right). \tag{12}$$

b) *The estimates* (11) *and* (12) *also hold true if we replace the sum by an integral over the same range and p by a continuous variable.*

Proof. We only prove part a), the proof of b) being similar.

The first inequality is obvious. Now notice that

$$\sum_{p\le\sqrt{x}} pe^{-c\sqrt{\log p}} \le \sum_{p\le x^{\frac{1}{2}(1-\epsilon)^2}} p + \sum_{x^{\frac{1}{2}(1-\epsilon)^2}<p\le\sqrt{x}} pe^{-c(\epsilon)\sqrt{\log x}} = \mathcal{O}\left(xe^{-c(\epsilon)\sqrt{\log x}}\right).$$

The proof of estimate (12) follows immediately from the observation

$$\sum_{\sqrt{x/r}<p\le\sqrt{x}} \frac{x}{p}e^{-c\sqrt{\log(x/p)}} \le xe^{-\frac{c}{\sqrt{2}}\sqrt{\log x}} \sum_{p\le\sqrt{x}}\frac{1}{p} = \mathcal{O}\left(xe^{-c(\epsilon)\sqrt{\log x}}\right),$$

where we used that $\sum_{p\le z} p^{-1} = \mathcal{O}(\log\log z)$. □

The sums in the next two lemmas arise on replacing π_{S_1} by Li in the second and third sum as appearing in (9).

Lemma 2. *Let $r\ge 1$ be an arbitrary fixed real number and S_1 any set of primes. Then*

$$\sum_{\substack{p\le\sqrt{x/r}\\ p\in S_1}} \mathrm{Li}(rp) = \pi_{S_1}(\sqrt{x/r})\mathrm{Li}(\sqrt{rx}) - \int_2^{\sqrt{rx}} \frac{\pi_{S_1}(t/r)}{\log t}\,dt.$$

Proof. We find that

$$\sum_{\substack{p\le\sqrt{x/r}\\ p\in S_1}} \mathrm{Li}(rp) = \sum_{\substack{p\le\sqrt{x/r}\\ p\in S_1}} \int_2^{rp}\frac{dt}{\log t} = \int_2^{\sqrt{rx}} \frac{A_{S_1}(t)}{\log t}\,dt,$$

where $A_{S_1}(t)=\#\{p\le\sqrt{x/r}\ :\ rp\ge t,\ p\in S_1\}$. The result easily follows on noting that $A_{S_1}(t)=\pi_{S_1}(\sqrt{x/r})-\pi_{S_1}(t/r)$ for $2\le t\le\sqrt{rx}$. □

Lemma 3. *Let $r > 1$ be an arbitrary fixed real number and S_1 any set of primes. Then*

$$\sum_{\substack{\sqrt{x/r}<p\le\sqrt{x}\\ p\in S_1}} \mathrm{Li}\left(\frac{x}{p}\right) = \pi_{S_1}(\sqrt{x})\mathrm{Li}(\sqrt{x}) - \pi_{S_1}(\sqrt{x/r})\mathrm{Li}(\sqrt{rx}) + \int\limits_{\sqrt{x}}^{\sqrt{rx}} \frac{\pi_{S_1}(x/t)}{\log t}\,dt$$

Proof. Note that

$$\sum_{\substack{\sqrt{x/r}<p\le\sqrt{x}\\ p\in S_1}} \mathrm{Li}\left(\frac{x}{p}\right) = \sum_{\substack{\sqrt{x/r}<p\le\sqrt{x}\\ p\in S_1}} \int\limits_{2}^{x/p} \frac{dt}{\log t} = \int\limits_{2}^{\sqrt{rx}} \frac{B_{S_1}(t)}{\log t}\,dt, \tag{13}$$

where $B_{S_1}(t) = \#\{\sqrt{x/r} < p \le \sqrt{x} \,:\, 2 \le t \le x/p,\ p \in S_1\}$. Clearly

$$B_{S_1}(t) = \begin{cases} \pi_{S_1}(\sqrt{x}) - \pi_{S_1}(\sqrt{x/r}) & \text{if } 2 \le t \le \sqrt{x};\\ \pi_{S_1}(x/t) - \pi_{S_1}(\sqrt{x/r}) & \text{if } \sqrt{x} < t \le \sqrt{rx}.\end{cases}$$

From this the result easily follows. □

Alternatively the lemmas 2 and 3 can be also proved by using partial integration and making an obvious linear transformation in the resulting integral. (The details are left to the interested reader.)

3.2. **Proof Theorem 2.** Recall that, for $r > 1$, $G_r(x)$ is defined by

$$G_r(x) = \frac{1}{2}\mathrm{Li}(\sqrt{x})^2 - \int_{2r}^{\sqrt{rx}} \frac{\mathrm{Li}(t/r)}{\log t}dt + \int_{\sqrt{x}}^{\sqrt{rx}} \frac{\mathrm{Li}(x/t)}{\log t}dt. \tag{14}$$

In our proof of Theorem 2 we will make use of the following observation.

Lemma 4. *We have* $G_r'(x) = \frac{1}{\log x}\left(\log\log\sqrt{rx} - \log\log\sqrt{x/r}\right)$.

Proof. The derivative of the first term on the right hand side of (14) is

$$\frac{d}{dx}\left(\frac{1}{2}\mathrm{Li}(\sqrt{x})^2\right) = \frac{\mathrm{Li}(\sqrt{x})}{\sqrt{x}\log x}.$$

The derivative of the first and the second integral on the right hand side of (14) equals

$$\frac{d}{dx}\left(\int_{2r}^{\sqrt{rx}} \frac{\mathrm{Li}(t/r)}{\log t}dt\right) = \frac{\mathrm{Li}(\sqrt{x/r})}{2\sqrt{x/r}\log\sqrt{rx}},$$

respectively

$$\frac{d}{dx}\left(\int_{\sqrt{x}}^{\sqrt{rx}} \frac{\mathrm{Li}(x/t)}{\log t}dt\right) = \frac{\mathrm{Li}(\sqrt{x/r})}{2\sqrt{x/r}\log\sqrt{rx}} - \frac{\mathrm{Li}(\sqrt{x})}{\sqrt{x}\log x} + \int_{\sqrt{x}}^{\sqrt{rx}} \frac{dt}{t\log t\log(x/t)}.$$

On adding them we get

$$G'{}_r(x) = \int_{\sqrt{x}}^{\sqrt{rx}} \frac{dt}{t\log t\log(x/t)}.$$

Note that

$$G_r'(x) = \frac{1}{\log x}\int_{\sqrt{x}}^{\sqrt{rx}} \frac{dt}{t\log t} + \frac{1}{\log x}\int_{\sqrt{x}}^{\sqrt{rx}} \frac{dt}{t\log(x/t)}.$$

By making a simple change of variable $t = x/v$ in the second integral on the right hand side above, we obtain

$$G_r'(x) = \frac{1}{\log x}\int_{\sqrt{x/r}}^{\sqrt{rx}} \frac{dv}{v\log v} = \frac{1}{\log x}\left(\log\log\sqrt{rx} - \log\log\sqrt{x/r}\right),$$

thus concluding the proof. □

Proof of Theorem 2. In (9) we replace every term $\pi_{S_2}(z)$ by the estimate given in (4) and invoke Lemma 1 a) to bound the resulting sums of error estimates giving rise to the asymptotic formula

$$\delta_2 D_r(x) = -\sum_{\substack{p\le\sqrt{x}\\ p\in S_1}} \mathrm{Li}(p) + \sum_{\substack{p\le\sqrt{x/r}\\ p\in S_1}} \mathrm{Li}(rp) + \sum_{\substack{\sqrt{x/r}<p\le\sqrt{x}\\ p\in S_1}} \mathrm{Li}\left(\frac{x}{p}\right) + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}}\right). \tag{15}$$

From (15), Lemmas 2, 3 and the observation that $\pi_{S_1}(z) = 0$ for $z < 2$ we infer that

$$\delta_2 D_r(x) = \int_2^{\sqrt{x}} \frac{\pi_{S_1}(t)}{\log t}dt - \int_{2r}^{\sqrt{rx}} \frac{\pi_{S_1}(t/r)}{\log t}dt + \int_{\sqrt{x}}^{\sqrt{rx}} \frac{\pi_{S_1}(x/t)}{\log t}dt + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}}\right).$$

By partial integration,

$$\int_2^{\sqrt{x}} \frac{\mathrm{Li}(t)}{\log t}dt = \frac{1}{2}\mathrm{Li}(\sqrt{x})^2. \tag{16}$$

Using this we see that if in the three integrals appearing in (15) we replace π_{S_1} by $\delta_1^{-1}\mathrm{Li}$ we obtain $G_r(x)/\delta_1$. Using Lemma 1 b) we estimate the sum of the errors made on making this replacement and conclude that as x tends to infinity we have

$$\delta_1\delta_2 D_r(x) = G_r(x) + \mathcal{O}\left(rxe^{-c(\epsilon)\sqrt{\log x}}\right). \tag{17}$$

Using Lemma 4 we notice that $G_r(x) - G_r(2r) = F_r(x)$ for $x \ge 2r$. Using some rough estimates on finds that $G_r(2r) = O(r)$ and hence we infer that $G_r(x) = F_r(x) + O(r)$. The proof is concluded on inserting this estimate in (17). □

4. Proof of Theorem 4

We will make use of the following lemma.

Lemma 5. *Let $\mathfrak{b} = \{b_j\}_{j=1}^{\infty}$ be a sequence of non-negative real numbers and $n \ge 2$ an arbitrary integer. We define*

$$N_{\mathfrak{b}}(x) := \sum_{j=1}^{n-1} b_j \int_2^x \frac{dt}{\log^j t}. \tag{18}$$

As x tends to infinity we have

$$N_{\mathfrak{b}}(x) = \sum_{k=1}^{n-1}\Big(\sum_{j=1}^{k} b_j\frac{(k-1)!}{(j-1)!}\Big)\frac{x}{\log^k x} + \mathcal{O}_n\left(\frac{xB_n}{\log^n x}\right),$$

where $B_n = \sum_{j=1}^{n} b_j$.

Proof. By partial integration one finds, with $n_j \ge 1$ an arbitrary integer,

$$\int_2^x \frac{dt}{\log^j t} = \sum_{m=1}^{n_j} \frac{(j+m-2)!}{(j-1)!} \frac{x}{\log^{j+m-1} x} + \mathcal{O}_{n_j}\left(\frac{x}{\log^{j+n_j} x}\right).$$

For $j = 1, \ldots, n-1$ we insert this in (18) and take, e.g., $n_j = n+1-j$. Rearranging terms then yields the result. □

Proof of Theorem 4. Let $|u| < 1$. Using the Taylor series

$$\log(1-u) = -\sum_{\ell=1}^{\infty} \frac{u^\ell}{\ell},$$

we conclude that

$$\log\left(\frac{1+u}{1-u}\right) = 2\sum_{\ell=1}^{\infty} \frac{u^{2\ell-1}}{2\ell-1}. \tag{19}$$

Define

$$E_m(u) = \log\left(\frac{1+u}{1-u}\right) - 2\sum_{\ell=1}^{m} \frac{u^{2\ell-1}}{2\ell-1}.$$

Note that

$$0 < E_m(u) = \frac{2u^{2m+1}}{2m+1} + \frac{2u^{2m+3}}{2m+3} + \cdots < 2\sum_{k=1}^{\infty} u^{2m+2k-1} = \frac{2u^{2m+1}}{1-u^2} \quad \text{for } 0 < u < 1. \tag{20}$$

Clearly

$$F_r'(x) = \frac{\log\log\sqrt{rx} - \log\log\sqrt{x/r}}{\log x} = \frac{1}{\log x}\left(\log\left(1+\frac{\log r}{\log x}\right) - \log\left(1-\frac{\log r}{\log x}\right)\right).$$

Recall that by assumption $x \ge 2r$. For those x we find by (19) the Taylor series

$$F_r'(x) = \frac{1}{\log x}\sum_{\ell=1}^{\infty} \frac{2}{2\ell-1}\left(\frac{\log r}{\log x}\right)^{2\ell-1}.$$

Using the definition of $E_m(u)$ it now follows that

$$F_r(x) = \int_{2r}^x F_r'(t)dt = \sum_{\ell=1}^{m} \frac{2}{2\ell-1}\log^{2\ell-1} r \int_{2r}^x \frac{dt}{\log^{2\ell} t} + \int_{2r}^x \frac{1}{\log t} E_m\left(\frac{\log r}{\log t}\right) dt. \tag{21}$$

From (20) we infer that, for $x \ge 2r$,

$$0 \le \int_{2r}^x \frac{1}{\log t} E_m\left(\frac{\log r}{\log t}\right) dt < \frac{2\log^{2m+1} r}{1-\left(\frac{\log r}{\log(2r)}\right)^2} \int_{2r}^x \frac{dt}{\log^{2m+2} t}.$$

Here, we note that

$$\frac{1}{1-\left(\frac{\log r}{\log(2r)}\right)^2} = \frac{\log^2(2r)}{(\log 2)\log(2r^2)} = \mathcal{O}\left(\log(2r)\right).$$

We conclude that

$$F_r(x) = \sum_{\ell=1}^{m} \frac{2}{2\ell-1}\log^{2\ell-1} r \int_{2r}^x \frac{dt}{\log^{2\ell} t} + \mathcal{O}_m\left(\frac{x\log(2r)\log^{2m+1} r}{\log^{2m+2} x}\right), \tag{22}$$

which can be rewritten as

$$F_r(x)=\sum_{\ell=1}^{m}\frac{2}{2\ell-1}\log^{2\ell-1}r\int_2^x\frac{dt}{\log^{2\ell}t}+\mathcal{O}_m\left(\frac{r}{\log(2r)}\right)+\mathcal{O}_m\left(\frac{x\log(2r)\log^{2m+1}r}{\log^{2m+2}x}\right),$$

where we used (6) to estimate $\int_2^{2r}dt/\log^{2\ell}t$. On noting that $r(\log(2r))^{-2}\log^{-2m-1}r$ is eventually increasing in r and $r\le x/2$ we see that

$$\frac{r}{\log(2r)}=\mathcal{O}_m\left(\frac{x\log(2r)\log^{2m+1}r}{\log^{2m+2}x}\right),$$

and therefore we have

$$F_r(x)=\sum_{\ell=1}^{m}\frac{2}{2\ell-1}\log^{2\ell-1}r\int_2^x\frac{dt}{\log^{2\ell}t}+\mathcal{O}_m\left(\frac{x\log(2r)\log^{2m+1}r}{\log^{2m+2}x}\right), \tag{23}$$

Lemma 5 applied with

$$b_j=\begin{cases}\frac{2}{j-1}\log^{j-1}r & \text{if } j \text{ is even;}\\ 0 & \text{otherwise,}\end{cases}$$

gives

$$\sum_{\ell=1}^{m}\frac{2}{2\ell-1}\log^{2\ell-1}r\int_2^x\frac{dt}{\log^{2\ell}t}=\sum_{k=1}^{2m}v_k(r)\frac{x}{\log^{k+1}x}+\mathcal{O}_m\left(\frac{x(\log r)\log^{2m-2}(2r)}{\log^{2m+2}x}\right), \tag{24}$$

where

$$v_k(r)=\sum_{j=1}^{[\frac{k+1}{2}]}\frac{k!}{(2j-1)!}\frac{2\log^{2j-1}r}{2j-1}=a_k(r).$$

On combining (24) with (23) the proof is then easily completed in case $n=2m+1$ is odd. (Observe that the error term in (24) is majorized by the one in (23).)

On noting that $a_{2m}(r)$ is an odd polynomial in $\log r$, we see that for all $r>1$ we have $a_{2m}(r)=\mathcal{O}_m((\log r)\log^{2m-2}(2r))$. Therefore

$$\frac{a_{2m}(r)x}{\log^{2m+1}x}=\mathcal{O}_m\Big(x\frac{(\log r)\log^{2m-2}(2r)}{\log^{2m+1}x}\Big),$$

and it follows from (24) that

$$\sum_{\ell=1}^{m}\frac{2}{2\ell-1}\log^{2\ell-1}r\int_2^x\frac{dt}{\log^{2\ell}t}=\sum_{k=1}^{2m-1}a_k(r)\frac{x}{\log^{k+1}x}+\mathcal{O}_m\left(\frac{x(\log r)\log^{2m-2}(2r)}{\log^{2m+1}x}\right). \tag{25}$$

On combining (25) with (23) the proof is then also completed in the remaining case where $n=2m$ is even. □

4.1. **Integrality of the coefficients of the polynomial $a_k(r)$.** Recall that

$$a_k(r)=\sum_{j=1}^{[(k+1)/2]}a_{k,j}\log^{2j-1}r,$$

with

$$a_{k,j}=\frac{k!2}{(2j-1)!(2j-1)}.$$

On being confronted with Table 1 the reader might wonder about the integrality of the coefficients $a_{k,j}$. The following result is easy to prove.

Proposition 1. *Define* $\mu(j) = \min\{k \geq 2j-1 : (2j-1)!(2j-1)|k!\}$.
a) *We have* $\mu(j) \leq 4j-2$ *with equality if and only if* $2j-1$ *is a prime number.*
b) *The coefficient* $a_{k,j}$ *is an integer if and only if* $k \geq \mu(j)$.
c) *Suppose that* $2j-1 = \prod_{p|2j-1} p^{e_p}$, *with all exponents* $e_p \leq p$. *Then*

$$\mu(j) = 2j-1+\max\{e_p p : p|2j-1\}.$$

5. Bias in the sense of Chebyshev

Let $\pi(x; d, a)$ denote the number of primes $p \leq x$ that satisfy $p \equiv a(\mathrm{mod}\ d)$. We restrict ourselves to the case where a and d are coprime, the cases where the residue class modulo d is said to be primitive. It is the only relevant case here as the non-primitive residue classes have only finitely many primes in them. It is a consequence of Legendre's theorem from 1837 that the primes are equidistributed over the primitive residue classes modulo d. Nevertheless, certain differences of the form $\pi(x; d, a_1) - \pi(x; d, a_2)$ are positive for many values of x (where "many" is best quantified using a logarithmic measure). This phenomenon was first observed and studied by Chebyshev who found that there is a strong bias for primes to be $\equiv 3(\mathrm{mod}\ 4)$ rather than $\equiv 1(\mathrm{mod}\ 4)$. For a survey see Granville and Martin [4].

Recently Ford and Sneed [3] and Xiangchang Meng [11] considered bias for products of two, respectively k primes, with $k \geq 2$ and fixed.

Problem 1. *Study the Chebyshev bias phenomenon for products of two proportional primes.*

Here especially the case where the modulus $d = 10$ is of relevance.

Acknowledgements. A large portion of this paper was written during the second author's stay in May 2016 at the Max Planck Institute for Mathematics (MPIM). She would like to thank Pieter Moree for inviting her and gratefully acknowledges support, hospitality as well as the excellent environment for collaboration at the MPIM. She is supported by the Austrian Science Fund (FWF): Project F5507-N26, which is a part of the Special Research Program "Quasi Monte Carlo Methods: Theory and Applications".

We would like to thank Florian Luca, Igor Shparlinski and Gérald Tenenbaum for helpful feedback. Furthermore we thank Alexandru Ciolan and Kate Kattegat for proofreading and Alex Weisse for his TeXnical help. The first author also acknowledges fruitful discussions with Yara Elias and Wadim Zudilin.

References

[1] A. Decker and P. Moree, Counting RSA-integers, *Result. Math.* **52** (2008), 35–39.
[2] D. Dummit, A. Granville and H. Kisilevsky, Big biases amongst products of two primes, arXiv: 1411.4594v1 (2014).
[3] K. Ford and J. Sneed, Chebyshev's bias for products of two primes, *Experiment. Math.* **19** (2010), 385–398.
[4] A. Granville and G. Martin, Prime number races, *Amer. Math. Monthly* **113** (2006), 1–33.
[5] J. von zur Gathen, CryptoSchool, Springer, Heidelberg, 2015.
[6] Y. Hashimoto, On asymptotic behavior of composite integers $n = pq$, *J. Math-for-Ind.* **1** (2009), 45–49.
[7] A. Hildebrand, On the number of prime factors of an integer. Ramanujan revisited (Urbana-Champaign, Ill., 1987), 167–185, Academic Press, Boston, MA, 1988.
[8] B. Justus, On integers with two prime factors, *Albanian J. Math.* **3** (2009), 189–197.
[9] E. Landau, Handbuch der Lehre von der Verteilung der Primzahlen, *Chelsea Publishing Co.,New York* (1953).
[10] D. Loebenberger, Grained integers and applications to cryptography, Dissertation, Universität Bonn, 2012. http://hss.ulb.uni-bonn.de/2012/2848/2848.pdf
[11] Xianchang Meng, Chebyshev's bias for products of k primes, arXiv:1606.04877.

[12] P.W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. Comput.* **26** (1997), 1484–1509.

Pieter Moree

Max-Planck-Institut für Mathematik, Vivatsgasse 7, D-53111 Bonn, Germany.

e-mail: moree@mpim-bonn.mpg.de

Sumaia Saad Eddin

Institute of Financial Mathematics and Applied Number Theory, Altenbergerstrasse 69, 4040 Linz, Austria.

e-mail: sumaia.saad_eddin@jku.at