Distributional Properties of the Largest Prime Factor

William D. Banks
Department of Mathematics, University of Missouri
Columbia, MO 65211 USA
bbanks@math.missouri.edu

Glyn Harman
Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
g.harman@rhul.ac.uk

Igor E. Shparlinski
Department of Computing, Macquarie University
Sydney, NSW 2109, Australia
igor@ics.mq.edu.au

October 8, 2005

Abstract

Let $P(n)$ denote the largest prime factor of an integer $n \geq 2$, and put $P(1) = 1$. In this paper, we study the distribution of the sequence $\{P(n) : n \geq 1\}$ over the set of congruence classes modulo an integer $q \geq 2$, and we study the same question for the sequence $\{P(p-1) : p \text{ is prime}\}$. We also give bounds for rational exponential sums involving $P(n)$. Finally, for an irrational number α, we show that the sequence $\{\alpha P(n) : n \geq 1\}$ is uniformly distributed modulo 1.

1 Introduction

For every positive integer n, let $P(n)$ denote the largest prime factor of n, with the usual convention that $P(1) = 1$. For an integer $q \geq 1$ and a real number z, we define $\mathbf{e}_q(z) = \mathbf{e}(z/q)$, where $e(z) = \exp(2\pi i z)$ as usual.

In Section 3 below, we consider the problem of bounding the function

$$\varrho(x; q, a) = \#\{n \leq x : P(n) \equiv a \pmod q\}.$$

In the case that q is fixed, this question has been previously considered by Ivić [11]. However, the approach of [11] does not appear to extend to the case where the modulus q is allowed to grow with the parameter x; this is mainly due to the fact that asymptotic formulas for the number of primes in arithmetic progressions are much less precise for growing moduli than those known for a fixed modulus.

We also remark that, in a recent work, Oon [13] has studied the distribution of $P(n)$ over the congruence classes of a fixed modulus q in the case that n itself belongs to an arithmetic progression (with a growing modulus).

In this paper, we use a similar approach to that of Ivić [11] and obtain new bounds that are nontrivial for a wide range of values of the parameter q; in particular, if q is not too large relative to x, we derive the expected asymptotic formula

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

with an explicit error term that is independent of a. On the other hand, we show that for $q \geq \exp\left(3\sqrt{\log x \log\log x}\,\right)$, this estimate is no longer correct (even by an order of magnitude).

In Section 4, we study the function

$$\varpi(x; q, a) = \#\{p \leq x : P(p-1) \equiv a \pmod q\},$$

where p varies over the set of prime numbers, and we derive the upper bound

$$\varpi(x; q, a) \ll \frac{\pi(x)}{\varphi(q)},$$

provided that $\log q \leq \log^{1/3} x$. Here, $\pi(x) = \#\{p \leq x\}$. We expect that the matching lower bound $\varpi(x; q, a) \gg \pi(x)/\varphi(q)$ also holds for such q, or perhaps even the stronger relation $\varpi(x; q, a) \sim \pi(x)/\varphi(q)$, but we have been unable to prove this. On the other hand, as in the case of $\rho(x; q, a)$, we expect that the behavior of $\varpi(x; q, a)$ changes for larger values of q. Unfortunately, the scarcity of the results about smooth shifted primes seems to be an obstacle to proving this.

In Section 5, we consider the related problem of bounding rational exponential sums of the form

$$S_{a,q}(x) = \sum_{n \le x} \mathbf{e}_q \left(aP(n) \right),$$

where the integers a and $q \ge 1$ are coprime. Our bounds are nontrivial if x is sufficiently large relative to q.

Finally, in Section 6, we bound the exponential sum

$$S_\alpha(x) = \sum_{n \le x} \mathbf{e}(\alpha P(n))$$

for a fixed irrational real number α. Our bound is nontrivial whenever x is sufficiently large, depending only on α, from which we deduce that the sequence $\{\alpha P(n) : n \ge 1\}$ is uniformly distributed modulo 1; this result is nicely reminiscent of the classical theorem of Vinogradov [15], which asserts that for a fixed irrational real number α, the sequence $\{\alpha p : p \text{ prime}\}$ is uniformly distributed modulo 1.

Our techniques are somewhat similar to those of [2, 3]. We expect that our underlying approach can be suitably modified to obtain nontrivial bounds for more general exponential and character sums involving the function $P(n)$.

Throughout the paper, the implied constants in the symbols "O", "$\gg$" and "$\ll$" are absolute (recall that the notations $U \ll V$ and $V \gg U$ are equivalent to the statement that $U = O(V)$ for positive functions U and V). We also use the symbol "o" with its usual meaning: the statement $U = o(V)$ is equivalent to $U/V \to 0$.

Throughout, p always denotes a prime number, $\log z$ denotes the natural logarithm of $z > 0$, and $\varphi(\cdot)$ and $\mu(\cdot)$ are the Euler and Möbius functions, respectively; we recall that $\mu(1) = 1$, $\mu(m) = 0$ if $m \ge 2$ is not squarefree, and $\mu(m) = (-1)^k$ if m is the product of k distinct primes.

2 Preliminary Estimates

As usual, we say that a positive integer n is *y-smooth* if and only if $P(n) \le y$. Let

$$\psi(x, y) = \#\{n \le x : n \text{ is } y\text{-smooth}\}.$$

The following estimate is a substantially relaxed and simplified version of the corollary to Theorem 3.1 of [4]; see also [10] and [14].

Lemma 1. *Let $u = (\log x)/(\log y)$, where $x \ge y > 0$. If $u \to \infty$ and $u \le y^{1/2}$, then the following estimate holds:*

$$\psi(x, y) = xu^{-u+o(u)}.$$

We remark that the condition $u \le y^{1/2}$ can be relaxed slightly, but this statement suffices for our purposes. To complement the estimate of Lemma 1, we also use the following bound, which holds for all $u \ge 1$ (see Theorem 1 of Chapter III.5 of [14]):

Lemma 2. *Let $u = (\log x)/(\log y)$, where $x \ge y > 0$. If $u \ge 1$, then the following bound holds:*

$$\psi(x, y) \ll x \exp(-u/2).$$

In what follows, we denote by $\mathcal{P}$ the set of all prime numbers, $\mathcal{P}[w, x]$ the set of primes p such that $w \le p \le x$, and for simplicity, we write $\mathcal{P}[x]$ for $\mathcal{P}[0, x]$. If the parameters $x \ge y > 0$ are fixed within a discussion, we also put $\mathcal{P}_m = \mathcal{P}[L_m, x/m]$ for all $m \ge 1$, where $L_m = \max\{y, P(m)\}$.

Lemma 3. *Let $x \ge y > 0$. For any two functions $h(k)$ and $f(k)$ satisfying $\max\{|h(k)|, |f(k)|\} \le 1$ for all positive integers k, we have*

$$\sum_{n \le x} h(P(n)) f(n) = \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} h(p) f(mp) + O(\psi(x, y)).$$

Proof. Denote by $\mathcal{N}$ the set of integers $n \le x$ with $P(n) \ge y$. Then

$$\sum_{n \le x} h(P(n)) f(n) = \sum_{n \in \mathcal{N}} h(P(n)) f(n) + O\left(\psi(x, y)\right). \tag{1}$$

Every integer $n \in \mathcal{N}$ has a unique representation of the form $n = mp$, where $p = P(n) \in \mathcal{P}_m$ and $m \le x/y$. Conversely, if $m \le x/y$ and $p \in \mathcal{P}_m$, then $n = mp$ lies in $\mathcal{N}$. Therefore,

$$\sum_{n \in \mathcal{N}} h(P(n)) f(n) = \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} h(P(mp)) f(mp) = \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} h(p) f(mp),$$

which together with (1) finishes the proof. □

As usual, we denote by $\pi(x; q, a)$ the number of primes $p \le x$ such that $p \equiv a \pmod q$. For a real number $x \ge 2$, write

$$\operatorname{li} x = \int_2^x \frac{d\,t}{\log t}.$$

We now recall the well known *Siegel–Walfisz theorem*; see Theorem 1.4.6 of [5] or, in an alternative form, Theorem 5 of Chapter II.8 of [14].

Lemma 4. *For every fixed number $A > 0$, there is a constant $B > 0$ such that for all $x \ge 2$ and all positive integers $q \le \log^A x$, the following bound holds:*

$$\max_{\gcd(a,q)=1} \left| \pi(x; q, a) - \frac{\operatorname{li} x}{\varphi(q)} \right| \ll x \exp\left(-B\sqrt{\log x}\right).$$

We also need the *Bombieri–Vinogradov theorem*; we refer the reader to Chapter 28 of [6], where it is given in a slightly different form from which the following statement can be derived by partial summation:

Lemma 5. *For every fixed number $A > 0$, there is a constant $B > 0$ such that for all $x \ge 2$, the following bound holds:*

$$\sum_{2 \le q \le x^{1/2} \log^{-B} x} \max_{\gcd(a,q)=1} \left| \pi(x; q, a) - \frac{\operatorname{li} x}{\varphi(q)} \right| \ll \frac{x}{\log^A x}.$$

Remark 1. *In [6], it is shown that one can take $B = A + 5$.*

In particular, for every fixed number $C > 0$, Lemma 5 implies that

$$\sum_{2 \le q \le X^{1/3}} \max_{\gcd(a,q)=1} \left| \pi(X; q, a) - \frac{\operatorname{li} X}{\varphi(q)} \right| \ll \frac{X}{\log^C X}, \qquad (2)$$

and this is the only form of Lemma 5 that is needed in the sequel.

The following two technical lemmas are needed for our study of $\varpi(x; q, a)$ in Section 4 below.

Lemma 6. *Uniformly for $q \le x$ the following bound holds:*

$$\max_{\gcd(b,q)=1} \sum_{\substack{m\le x\\ \gcd(m-b,q)=1}} \frac{1}{\varphi(m)} \ll \frac{\varphi(q)}{q} \log x.$$

Proof. We start with the identity:

$$\begin{aligned}
\sum_{\substack{m\le x\\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} &= \sum_{m\le x} \sum_{c \mid \gcd(m-b,q)} \mu(c) \sum_{d\mid m} \frac{\mu^2(d)}{\varphi(d)} \\
&= \sum_{d\le x} \frac{\mu^2(d)}{\varphi(d)} \sum_{c\mid q} \mu(c) \sum_{\substack{m\le x,\ d\mid m\\ c\mid m-b}} 1 \\
&= \sum_{d\le x} \frac{\mu^2(d)}{\varphi(d)} \sum_{c\mid q} \mu(c) \sum_{\substack{n\le x/d\\ dn\equiv b \pmod c}} 1.
\end{aligned}$$

Note that the last sum is empty unless $\gcd(c,d)=1$, in which case we have

$$\sum_{\substack{n\le x/d\\ dn\equiv b \pmod c}} 1 = \frac{x}{cd} + O(1).$$

It follows that

$$\begin{aligned}
\sum_{\substack{m\le x\\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} &= \sum_{d\le x} \frac{\mu^2(d)}{\varphi(d)} \sum_{\substack{c\mid q\\ \gcd(c,d)=1}} \mu(c) \left(\frac{x}{cd} + O(1)\right) \\
&= x \sum_{d\le x} \frac{\mu^2(d)}{d\varphi(d)} \sum_{\substack{c\mid q\\ \gcd(c,d)=1}} \frac{\mu(c)}{c} + O\left(2^{\omega(q)} \log x\right),
\end{aligned}$$

where $\omega(q)$ is the number of distinct prime divisors of q; here, we have used the result of Landau that (see, for example, [12]) that

$$\sum_{d\le x} \frac{1}{\varphi(d)} \ll \log x$$

(see [12] for a more precise statement). Now,

$$\sum_{\substack{c \mid q\\ \gcd(c,d)=1}} \frac{\mu(c)}{c} = \prod_{\substack{p \mid q\\ p \nmid d}} \left(1 - \frac{1}{p}\right) = \frac{\varphi(q)}{q} \frac{\gcd(d,q)}{\varphi(\gcd(d,q))},$$

and therefore,

$$\sum_{\substack{m \le x\\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} = \frac{\varphi(q)}{q} x \sum_{d \le x} \frac{\mu^2(d)}{d\varphi(d)} \frac{\gcd(d,q)}{\varphi(\gcd(d,q))} + O\left(2^{\omega(q)} \log x\right).$$

Noting that

$$\sum_{d \le x} \frac{\mu^2(d)}{d\varphi(d)} \frac{\gcd(d,q)}{\varphi(\gcd(d,q))} \ll 1,$$

and for all $q \le x$,

$$2^{\omega(q)} \log x \ll \frac{\varphi(q)}{q} x,$$

we obtain that

$$\sum_{\substack{m \le x\\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} \ll \frac{\varphi(q)}{q} x,$$

uniformly for $q \le x$ and b coprime to q. The result now follows by partial summation. □

Lemma 7. *Uniformly for* $\exp(\log^{1/5} x) \le y \le x^{1/2}$, $q \le \exp(\log^{6/7} y)$, *the following bound holds:*

$$\max_{\gcd(b,q)=1} \sum_{\substack{m \le x\\ P(m) \le y\\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} \ll \frac{\varphi(q)}{q} \psi(x,y).$$

Proof. Write

$$\begin{aligned} \sum_{\substack{m \le x\\ P(m) \le y\\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} &= \sum_{\substack{m \le x\\ P(m) \le y}} \frac{m}{\varphi(m)} \sum_{c \mid \gcd(m-b,q)} \mu(c) \\ &= \sum_{c \mid q} \mu(c) \sum_{\substack{m \le x\\ P(m) \le y\\ m \equiv b \pmod{c}}} \frac{m}{\varphi(m)}. \end{aligned}$$

Using Theorem 1 of [1] together with the well known estimate

$$\psi(x,y) = \rho(u)\, x \left(1 + O\left(\frac{\log u}{\log y}\right)\right),$$

where $\rho(\cdot)$ is the Dickman function and $u = (\log x)/(\log y)$, we obtain the estimate

$$\sum_{\substack{m \le x \\ P(m) \le y \\ m \equiv b \pmod c}} \frac{m}{\varphi(m)} = \frac{\zeta_c(2)\zeta_c(3)}{\zeta_c(6)} \frac{\psi(x,y)}{c} \left(1 + O\left(\log^{-1/35} x\right)\right),$$

which is uniform in all parameters subject to the specified constraints. Here, $\zeta_c(s)$ is the partial zeta-function that is defined for $\Re(s) > 1$ by

$$\zeta_c(s) = \prod_{p \nmid c} (1 - p^{-s})^{-1}.$$

Consequently,

$$\sum_{\substack{m \le x \\ P(m) \le y \\ \gcd(m-b,q)=1}} \frac{m}{\varphi(m)} = \psi(x,y) \sum_{c \,|\, q} \frac{\mu(c)}{c} \frac{\zeta_c(2)\zeta_c(3)}{\zeta_c(6)} + O\left(\frac{\psi(x,y)}{\log^{1/35} x} \sum_{c \,|\, q} \frac{\mu^2(c)}{c}\right).$$

Since

$$\sum_{c \,|\, q} \frac{\mu^2(c)}{c} = \prod_{p \,|\, q} \left(1 + \frac{1}{p}\right) \ll \frac{q}{\varphi(q)} \ll \frac{\varphi(q)}{q} \log^{1/36} x,$$

the error term above is of size $o\left(\varphi(q)\psi(x,y)/q\right)$. For the main term, we observe that

$$\begin{aligned}
\sum_{c \,|\, q} \frac{\mu(c)}{c} \frac{\zeta_c(2)\zeta_c(3)}{\zeta_c(6)} &= \frac{\zeta(2)\zeta(3)}{\zeta(6)} \sum_{c \,|\, q} \frac{\mu(c)}{c} \prod_{p \,|\, c} \frac{(1-p^{-2})(1-p^{-3})}{(1-p^{-6})} \\
&\ll \prod_{p \,|\, q} \left(1 - \frac{1}{p} \frac{(1-p^{-2})(1-p^{-3})}{(1-p^{-6})}\right) \\
&= \frac{\varphi(q)}{q} \prod_{p \,|\, q} \left(1 + \frac{1}{p^3 - 2p^2 + 2p - 1}\right) \ll \frac{\varphi(q)}{q},
\end{aligned}$$

and the result follows. □

One of our principal tools is the following bound for exponential sums over prime numbers, which follows immediately from Chapter 25 of [6] by partial summation (see also [2, 3]):

Lemma 8. *Let $\alpha \in \mathbb{R}$, and suppose that there are integers a, q with $q \geq 1$, $\gcd(a, q) = 1$, and*

$$\left| \alpha - \frac{a}{q} \right| < \frac{1}{q^2}.$$

Then, for all $x \geq 2$, the following bound holds:

$$\left| \sum_{p \in \mathcal{P}[x]} \mathbf{e}(\alpha p) \right| \ll x \left(q^{-1/2} + x^{-1/5} + q^{1/2} x^{-1/2} \right) \log^3 x.$$

Finally, we also need the following "major arc" bound, which can be deduced from the bound on page 147 in Chapter 26 of [6] by partial summation:

Lemma 9. *For every fixed number $A > 0$, there is a constant $B > 0$ with the following property. Let $x \geq 2$, and suppose that*

$$\alpha = \frac{a}{q} + \beta,$$

where a, q are coprime integers, and

$$1 \leq q \leq \log^A x, \qquad |\beta| < \frac{\log^A x}{x}.$$

Then

$$\sum_{p \in \mathcal{P}[x]} \mathbf{e}(\alpha p) = \frac{\mu(q)}{\varphi(q)} \sum_{n \leq x} \frac{\mathbf{e}(n\beta)}{\log n} + O\left(x \exp\left(-B\sqrt{\log x} \right) \right).$$

In particular,

$$\left| \sum_{p \in \mathcal{P}[x]} \mathbf{e}(\alpha p) \right| \ll \frac{x}{\varphi(q) \log x}.$$

3 Distribution of $P(n)$ in Congruence Classes

Theorem 1. *For every fixed number $\Delta > 0$, there is a constant $c > 0$ such that for any positive integer q, the following bound holds:*

$$\max_{\gcd(a,q)=1} \left| \varrho(x;q,a) - \frac{x}{\varphi(q)} \right| \ll x \left(x^{-q^{-\Delta}} + \exp(-c \log^{1/3} x) \right).$$

Proof. Throughout the proof, let a be fixed with $\gcd(a,q) = 1$. Consider the function $h(k)$ defined by

$$h(k) = \begin{cases} 1 & \text{if } k \equiv a \pmod q, \\ 0 & \text{otherwise.} \end{cases}$$

Put

$$y = \exp\left(\tfrac{1}{2} q^{\Delta}\right) \qquad \text{and} \qquad u = \frac{\log x}{\log y} = 2q^{-\Delta} \log x.$$

By Lemmas 2 and 3, we have

$$\varrho(x;q,a) = \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} h(p) + O\left(x \exp(-u/2)\right). \tag{3}$$

For any m with $mL_m \le x$, we have

$$\sum_{p \in \mathcal{P}_m} h(p) = \pi(x/m;q,a) - \pi(L_m;q,a) + O(1),$$

and the sum is empty otherwise. We observe that the error term in the bound in Lemma 4 is a monotonically increasing function of x; thus, for all positive integers m with $x/m \ge L_m \ge y$, since $q \le 2 \log^{1/\Delta} y$, the estimate

$$\sum_{p \in \mathcal{P}_m} h(p) = \frac{1}{\varphi(q)} \left(\operatorname{li}(x/m) - \operatorname{li} L_m \right) + O\left(xm^{-1} \exp\left(-c_1 \sqrt{\log(x/m)} \right) \right)$$

holds for some constant $c_1 > 0$ depending only on Δ. Therefore, by (3) we obtain that

$$\varrho(x;q,a) = \frac{1}{\varphi(q)} \sum_{m \le x/y} \left(\operatorname{li}(x/m) - \operatorname{li} L_m \right) + O\left(x^{1-q^{-\Delta}} + R \right),$$

where

$$R = x \sum_{\substack{m \le x/y \\ mL_m \le x}} m^{-1} \exp\left(-c_1 \sqrt{\log(x/m)}\right).$$

The same arguments applied with $h(k) = 1$ lead to the identity

$$\lfloor x \rfloor = \sum_{n \le x} 1 = \sum_{m \le x/y} (\operatorname{li}(x/m) - \operatorname{li} L_m) + O\left(x^{1-q^{-\Delta}} + R\right).$$

Hence,

$$\varrho(x; q, a) = \frac{x}{\varphi(q)} + O\left(x^{1-q^{-\Delta}} + R\right). \tag{4}$$

To estimate R, we put $L = \lceil \log(x/y) \rceil$, and derive that

$$\begin{aligned} R &\le x \sum_{j=1}^{L} \sum_{\substack{e^{j-1} \le m < e^j \\ P(m) \le x/m}} m^{-1} \exp\left(-c_1 \sqrt{\log(x/m)}\right) \\ &\ll x \sum_{j=1}^{L} \exp\left(-c_1 \sqrt{\log(x/e^j)}\right) \exp\left(-\frac{\log(e^j)}{2\log(x/e^j)}\right), \end{aligned}$$

where we have used Lemma 2 in the last step. Now we have the inequality

$$c_1 \sqrt{\log(x/M)} + \frac{\log M}{2\log(x/M)} \ge c_2 \log^{1/3} x$$

for some absolute constant $c_2 > 0$ and all $x > M > 0$, as is readily verified by considering the cases $M < x \exp\left(\log^{-2/3} x\right)$ and $M \ge x \exp\left(\log^{-2/3} x\right)$ separately. Thus, it follows that

$$R \ll xL \exp\left(-c_2 \log^{1/3} x\right) \ll x \exp\left(-c \log^{1/3} x\right),$$

for some constant $c > 0$. Combining this result with (4) finishes the proof. □

It is clear that Theorem 1 is only non-trivial when $q \le \log^K x$ for a fixed constant $K > 0$, which is to be expected given our limited knowledge concerning primes in arithmetic progressions. Of course, much better results for primes in arithmetic progressions are known "on average" as the modulus q varies over all values up to $\sqrt{x}/\log^B x$, as evidenced by Lemma 5. On the other hand, Theorem 1 cannot be extended to such a wide range, since the largest prime divisor $P(n)$ often takes very small values; this limitation is encapsulated in the following result:

Theorem 2. *For every sufficiently large number x, there exists an integer a such that the lower bound*

$$\varrho(x;q,a) > \frac{x}{\varphi(q)^{1/2}}$$

holds for every modulus $q \ge \exp\bigl(3\sqrt{\log x \log\log x}\,\bigr)$.

Proof. Put

$$v = \sqrt{\frac{2\log x}{\log\log x}},$$

and let a be the prime number lying closest to $x^{1/v}$; then $a = (1+o(1))x^{1/v}$. Consider the set of products $n = ma$, where m runs over all positive integers $m \le x/a$ that are $(a-1)$-smooth. Clearly, each integer n is counted by $\varrho(x;q,a)$ for every modulus q; therefore,

$$\varrho(x;q,a) \ge \psi(x/a, a-1).$$

Since $\log(x/a)/\log a = v + o(v)$, using Lemma 1 we derive that

$$\begin{aligned} \varrho(x;q,a) &\ge \frac{x}{av^{v+o(v)}} = x\exp\bigl(-v^{-1}\log x - (1+o(1))v\log v\bigr) \\ &= x\exp\left(-\sqrt{(2+o(1))\log x \log\log x}\right), \end{aligned}$$

and the result follows. □

In view of the lower bound of Theorem 2, the following analogue of the Bombieri–Vinogradov theorem, which at first glance appears somewhat weak, is nevertheless the best result possible in our situation:

Theorem 3. *For every fixed number $B > 0$ and all $x \ge 2$, we have*

$$\sum_{2\le q\le \exp(\sqrt{\log x}\,)} \max_{\gcd(a,q)=1} \left| \varrho(x;q,a) - \frac{x}{\varphi(q)} \right| \ll \frac{x}{\log^B x}.$$

Proof. Put $C = 2B+2$, and let us define

$$y = \exp\left(3\sqrt{\log x}\right) \qquad \text{and} \qquad u = \frac{\log x}{\log y} = \frac{\sqrt{\log x}}{3}.$$

Arguing as in the proof of Theorem 1, but applying Lemma 1 rather than Lemma 2, we are led to the estimate:

$$\sum_{2\le q\le \exp(\sqrt{\log x}\,)} \max_{\gcd(a,q)=1} \left|\varrho(x;q,a) - \frac{x}{\varphi(q)}\right| \ll x\exp\left(\sqrt{\log x}\,\right) u^{-u+o(u)} + R$$

where

$$R = x \sum_{\substack{m\le x/y\\ mL_m\le x}} m^{-1}\log^{-C}(x/m).$$

Here, we have applied (2) with $X = x/m$; note that our choice of y guarantees that $q \le X^{1/3}$ for all q in the stated range. Trivially, we have

$$R \le x \sum_{m\le x/y} m^{-1}\log^{-C}(x/m) \le x\log^{-C} y \sum_{m\le x/y} m^{-1} \ll x\log^{2-C} y,$$

and the result follows from our choices of C, y and u. □

4 Distribution of $P(p-1)$ Modulo q

Recall that

$$\varpi(x;q,a) = \#\{p\le x : P(p-1)\equiv a \pmod q\}.$$

Theorem 4. *For all $q \le \exp(\log^{1/3} x)$, the following bound holds:*

$$\max_{\gcd(a,q)=1} \varpi(x;q,a) \ll \frac{\pi(x)}{\varphi(q)}.$$

Proof. Throughout the proof, let a be fixed with $\gcd(a,q)=1$, and put

$$y = \exp(\log^{2/5} x) \qquad \text{and} \qquad u = \frac{\log x}{\log y} = \log^{3/5} x,$$

where x is a large real number. Note that by Lemma 1, we have

$$\psi(x,y) = x\exp\left(-(0.6+o(1))(\log x)^{3/5}\log\log x\right) \ll \frac{\pi(x)}{\varphi(q)}. \tag{5}$$

Let $h(k)$ and $f(k)$ be the functions given by

$$h(k) = \begin{cases} 1 & \text{if } k\equiv a \pmod q,\\ 0 & \text{otherwise,}\end{cases}$$

and

$$f(k) = \begin{cases} 1 & \text{if } k+1 \text{ is prime,} \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 3 and the bound (5), we have

$$\begin{aligned} \varpi(x;q,a) &= \sum_{n \le x} h(P(n)) f(n) + O(1) \\ &= \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} h(p) f(mp) + O\left(\frac{\pi(x)}{\varphi(q)}\right). \end{aligned}$$

For each integer m, observe that

$$\sum_{p \in \mathcal{P}_m} h(p) f(mp) = \#\{p \in \mathcal{P}_m : p \equiv a \pmod q \text{ and } mp+1 \text{ is prime}\}. \quad (6)$$

Note that the sum is empty unless $mL_m \le x$. If $\gcd(am+1,q) = d > 1$, then writing $p = qt + a$, we see that $mp + 1 = mqt + am + 1$ is divisible by d; this shows that the right side of (6) is either 0 or 1 for every such m. Since $x/y \ll \pi(x)/\varphi(q)$ by our choice of y, it follows that

$$\varpi(x;q,a) = \sum_{\substack{m \le x/y \\ P(m) \le x/m \\ \gcd(am+1,q)=1}} \sum_{p \in \mathcal{P}_m} h(p) f(mp) + O\left(\frac{\pi(x)}{\varphi(q)}\right).$$

If $m \le x/y$ and $\gcd(am+1,q) = 1$, we can apply a standard sieve to bound the right side of (6) (see, for example, Corollary 2.4.1 of [8]; note that $q < x/m$ by our choice of q), and for such m we obtain that

$$\begin{aligned} \sum_{p \in \mathcal{P}_m} h(p) f(mp) &\ll \frac{x/m}{\varphi(q) \log^2(x/mq)} \prod_{p \mid mq} \left(1 - \frac{1}{p}\right)^{-1} \\ &\le \frac{q}{\varphi(q)^2} \cdot \frac{x}{\varphi(m) \log^2(x/mq)}. \end{aligned}$$

Thus, to complete the proof, it suffices to show that

$$T = \sum_{\substack{m \le x/y \\ P(m) \le x/m \\ \gcd(am+1,q)=1}} \frac{x}{\varphi(m) \log^2(x/mq)} \ll \frac{\varphi(q)}{q} \pi(x).$$

Splitting the sum T into two pieces, we see that $T \ll T_1 + T_2$, where

$$\begin{aligned} T_1 &= \frac{x}{\log^2 x} \sum_{\substack{m \le x^{2/3} \\ \gcd(am+1,q)=1}} \frac{1}{\varphi(m)}, \\ T_2 &= \sum_{\substack{x^{2/3} < m \le x/y \\ P(m) \le x/m \\ \gcd(am+1,q)=1}} \frac{x}{\varphi(m) \log^2(x/mq)}. \end{aligned}$$

For T_1, we apply Lemma 6 with $b \equiv -a^{-1} \pmod q$, which gives

$$T_1 \ll \frac{x}{\log^2 x} \frac{\varphi(q)}{q} \log x \ll \frac{\varphi(q)}{q} \pi(x).$$

To estimate T_2, put $M = \left\lfloor \frac{2}{3} \log x \right\rfloor$ and $L = \lceil \log(x/y) \rceil$; then by Lemmas 2 and 7 (again with $b \equiv -a^{-1} \pmod q$), we have

$$\begin{aligned} T_2 &\ll x \sum_{j=M}^{L} \frac{1}{e^j \log^2(x/e^j q)} \sum_{\substack{e^{j-1} \le m < e^j \\ P(m) \le x/e^{j-1} \\ \gcd(am+1,q)=1}} \frac{m}{\varphi(m)} \\ &\ll \frac{\varphi(q)}{q} x \sum_{j=M}^{L} \frac{1}{\log^2(x/e^j q)} \exp\left(-\frac{\log(e^j)}{2\log(x/e^{j-1})}\right). \end{aligned}$$

Since $q^2 = o(y)$, we see that $x/e^j q \ge (x/e^{j-1})^{1/2}$ for all $j \le L$ if x is large enough. Therefore,

$$\begin{aligned} T_2 &\ll \frac{\varphi(q)}{q} x \sum_{j=M}^{L} \frac{1}{\log^2(x/e^{j-1})} \exp\left(\frac{\log(x/e^{j-1}) - \log x}{2\log(x/e^{j-1})}\right) \\ &\ll \frac{\varphi(q)}{q} \frac{x}{\log^2 x} \sum_{j=M}^{L} \frac{\log^2 x}{\log^2(x/e^{j-1})} \exp\left(-\frac{\log x}{2\log(x/e^{j-1})}\right) \\ &\ll \frac{\varphi(q)}{q} \frac{xL}{\log^2 x} \ll \frac{\varphi(q)}{q} \frac{x}{\log x} \ll \frac{\varphi(q)}{q} \pi(x), \end{aligned}$$

and the proof is complete. □

5 Rational Exponential Sums with $P(n)$

We now show that arguments from [2, 3] can be used to estimate rational exponential sums with $P(n)$.

Theorem 5. *For any integer $q \ge 2$, the bound*

$$\max_{\gcd(a,q)=1} |S_{a,q}(x)| \ll x \left(v^{-2v/5+o(v)} + q^{-1/2}\log^4 x\right)$$

holds with $v = (\log x)/(\log q)$.

Proof. Without loss of generality, we can also assume that $q \ge \log^8 x$ since the bound is trivial otherwise. Throughout the proof, fix a with $\gcd(a,q) = 1$. We define $y = q^{5/2}$ and remark that

$$u = \frac{\log x}{\log y} = \frac{2v}{5} \le \log x \le y^{1/2},$$

thus we can apply Lemma 1. By Lemma 3, applied with $h(k) = \mathbf{e}_q(ak)$, we see that

$$S_{a,q}(x) = \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} \mathbf{e}_q(ap) + O\left(xu^{-u+o(u)}\right), \tag{7}$$

where as before $\mathcal{P}_m = \mathcal{P}[L_m, x/m]$ and $L_m = \max\{y, P(m)\}$. Write

$$\sum_{p \in \mathcal{P}[L_m, x/m]} \mathbf{e}_q(ap) = \sum_{p \in \mathcal{P}[x/m]} \mathbf{e}_q(ap) - \sum_{p \in \mathcal{P}[L_m - 1]} \mathbf{e}_q(ap).$$

Now, by Lemma 8, we have for all positive integers $m \le x/y$:

$$\begin{aligned}\sum_{p \in \mathcal{P}_m} \mathbf{e}_q(ap) &\ll \frac{x}{m}\left(q^{-1/2} + x^{-1/5}m^{1/5} + q^{1/2}x^{-1/2}m^{1/2}\right)\log^3 x \\ &\ll \frac{x}{m}\left(q^{-1/2} + y^{-1/5} + q^{1/2}y^{-1/2}\right)\log^3 x.\end{aligned}$$

Recalling the definition of y, we see that the first term always dominates; therefore,

$$\sum_{p \in \mathcal{P}_m} \mathbf{e}_q(ap) \ll \frac{x \log^3 x}{mq^{1/2}}.$$

Consequently,

$$\sum_{m\le x/y}\left|\sum_{p\in\mathcal{P}[L_m,x/m]}\mathbf{e}_q(ap)\right| \ll \frac{x\log^3 x}{q^{1/2}}\sum_{m\le x/y}\frac{1}{m}=\frac{x\log^4 x}{q^{1/2}},$$

which together with (7) finishes the proof. □

As we have remarked earlier, the bound of Theorem 5 is trivial when $q\le\log^8 x$. Fortunately, the result of Theorem 1 can be used to provide a bound on $S_{a,q}(x)$ that is nontrivial for all moduli $q\le\log^A x$, where $A>0$ is any fixed constant.

Theorem 6. *For every fixed number $\Delta>0$, there is a constant $c>0$ such that for any positive integer q, the following bound holds:*

$$\max_{\gcd(a,q)=1}|S_{a,q}(x)|\ll x\left(|\mu(q)|\varphi(q)^{-1}+x^{-q^{-\Delta}}+\exp(-c\log^{1/3}x)\right).$$

Proof. Throughout the proof, fix a with $\gcd(a,q)=1$. Applying Theorem 1 with $\Delta/2$ instead of Δ, we obtain that for some constant $c_1>0$,

$$\begin{aligned}S_{a,q}(x) &= \sum_{\substack{b=1\\ \gcd(b,q)=1}}^{q}\varrho(x;q,b)\,\mathbf{e}_q(ab)\\ &= \frac{x}{\varphi(q)}\sum_{\substack{b=1\\ \gcd(b,q)=1}}^{q}\mathbf{e}_q(ab)+O\left(x(qx^{-q^{-\Delta/2}}+q\exp(-c_1\log^{1/3}x)\right)\\ &= \frac{x}{\varphi(q)}\sum_{\substack{b=1\\ \gcd(b,q)=1}}^{q}\mathbf{e}_q(b)+O\left(x(qx^{-q^{-\Delta/2}}+q\exp(-c_1\log^{1/3}x)\right).\end{aligned}$$

The sum over b is the well-known *Ramanujan sum*, which evaluates to

$$\sum_{\substack{b=1\\ \gcd(b,q)=1}}^{q}\mathbf{e}_q(b)=\mu(q);$$

see, for example, Theorem 272 in [7]. If $q\ge\log^{1/\Delta}x$, the bound of the theorem is trivial, while for $q<\log^{1/\Delta}x$, we have

$$qx^{-q^{-\Delta/2}}\le x^{-q^{-\Delta}} \qquad\text{and}\qquad q\exp(-c_1\log^{1/3}x)\ll\exp(-0.5\,c_1\log^{1/3}x).$$

The result follows. □

We remark that if q is squarefree, then for $q < \log^{1/\Delta} x$ the last term never dominates the first one, hence the bound of Theorem 6 takes the form

$$\max_{\gcd(a,q)=1} |S_{a,q}(x)| \ll x \left(\varphi(q)^{-1} + x^{-q^{-\Delta}} \right).$$

On the other hand, if q is not squarefree, then the first term simply disappears, and the bound of Corollary 6 takes the form

$$\max_{\gcd(a,q)=1} |S_{a,q}(x)| \ll x \left(x^{-q^{-\Delta}} + \exp(-c \log^{1/3} x) \right).$$

6 Distribution of $\alpha P(n)$ Modulo 1

Our goal here is to replace a/q in the previous section by an arbitrary real number α and obtain a bound for $S_\alpha(x)$ which implies that the sequence $\{\alpha P(n) : n \ge 1\}$ is uniformly distributed modulo 1 when α is irrational.

Theorem 7. *Let $x \ge 2$ and $\alpha \in \mathbb{R}$. Let $\{a_j/q_j : j = 1, 2, \ldots\}$ be the sequence of convergents in the continued fraction expansion of α, and put*

$$g = \max \left\{ q_j : q_j < \exp\left(\sqrt{\log x}\right) \right\}.$$

Then

$$\left| \sum_{n \le x} \mathbf{e}(\alpha P(n)) \right| \ll x g^{-1/3}.$$

Proof. Let $y = \max\left\{ \exp(3 \log^{1/2} x),\, x^{g^{-1/2}} \right\}$. As before, we have

$$\sum_{n \le x} \mathbf{e}(\alpha P(n)) = \sum_{m \le x/y} \sum_{p \in \mathcal{P}_m} \mathbf{e}(\alpha p) + O\left(x u^{-u + o(u)} \right),$$

where

$$u = \frac{\log x}{\log y} = \min\left(\frac{\log^{1/2} x}{3}, g^{1/2} \right).$$

Note that $u^{-u+o(u)} \ll g^{-1/3}$ since $g \le \exp(\log^{1/2} x)$.

Suppose first that $g \ge \log^{24} x$. Then, by Lemma 8, we have

$$\sum_{p \in \mathcal{P}_m} \mathbf{e}(\alpha p) \ll \frac{x}{m} \left(g^{-1/2} + (m/x)^{1/5} + (gm/x)^{1/2} \right) \log^3(x/m).$$

Hence

$$\sum_{m\le x/y}\sum_{p\in\mathcal{P}_m}\mathbf{e}(\alpha p)\ll\log^3 x\left(\frac{x}{g^{1/2}}\log x+\frac{x}{y^{1/5}}+x\left(\frac{g}{y}\right)^{1/2}\right)\ll xg^{-1/3},$$

by our choice of parameters.

Now suppose that $g<\log^{24}x$. For each m, let

$$r_m=\max\left\{q_j:q_j\le\frac{x}{m\log^{24}x}\right\}.$$

If $r_m>\log^{24}x$, then

$$\sum_{p\in P_m}\mathbf{e}(\alpha p)\ll\frac{x}{m\log^9 x}$$

by Lemma 8; summing this bound over all such m gives a bound of order $O\left(x\log^{-8}x\right)=O(xg^{-1/3})$. On the other hand, if $r_m<\log^{24}x$, then $r_m=g$ (by the properties of convergents in a continued fraction). We can therefore use Lemma 9, which gives

$$\left|\sum_{p\in P_m}\mathbf{e}(\alpha p)\right|\ll\frac{x/m}{\varphi(g)\log(x/m)}.\tag{8}$$

Summing (8) over all possible m gives the upper bound

$$\left|\sum_{n\le x}\mathbf{e}(\alpha P(n))\right|\ll\frac{x\log x}{\varphi(g)\log y}+x(\log x)\exp\left(-C\log^{1/4}x\right)\ll xg^{-1/3},$$

by our choice of parameters (note that $\log y\ge(\log x)/g^{1/2}$). This completes the proof. □

Remark 2. *It is possible to improve $g^{1/3}$ in the bound of Theorem 7 to $g^{1/2-\varepsilon}$ for any fixed $\varepsilon>0$.*

Remark 3. *If α is irrational, then there are infinitely many convergents in its continued fraction; thus, the parameter g in Theorem 7 tends to infinity with x.*

Let $\{\vartheta\}$ denote the fractional part of the real number ϑ. We recall that the discrepancy $D(x)$ of an arbitrary sequence $\{\vartheta_n : n \ge 1\}$ is defined as

$$D(x) = \sup_{0 \le \gamma \le 1} \left| N_\gamma(x) - \gamma x \right|.$$

where $N_\gamma(x)$ is the counting function

$$N_\gamma(x) = \#\{n \le x : \{\vartheta_n\} \le \gamma\}.$$

The sequence is said to be uniformly distributed modulo 1 if

$$\lim_{x \to \infty} \frac{D(x)}{x} = 0.$$

Corollary 1. *If α is irrational, then the sequence $\{\alpha P(n) : n \ge 1\}$ is uniformly distributed modulo 1.*

Proof. By Weyl's criterion (see Theorem 5.6 in [9]), we need only show that

$$\sum_{n \le x} \mathbf{e}(\alpha h P(n)) = o(x)$$

for every integer $h \ge 1$. Since α is irrational, αh is also irrational, and the result follows immediately from Theorem 7 in view of our remark above that $g \to \infty$ as $x \to \infty$. □

Remark 4. *Unfortunately, there is no hope of getting an explicit discrepancy bound in Theorem 1 unless one assumes an appropriate condition for α, since one can always "manufacture" real numbers α for which the discrepancy decreases at an arbitrarily slow rate.*

We recall that α is called a *Liouville number* if

$$\limsup_{q \to \infty} \frac{\log \|\alpha q\|^{-1}}{\log q} = \infty.$$

In the case that α is *not* a Liouville number (which is the case for almost all real α), we have the following result:

Corollary 2. *Let $D(x)$ be the discrepancy of the sequence $\{\alpha P(n) : n \le x\}$. Then, provided that α is not a Liouville number, we have*

$$D(x) \ll x \exp\left(-\sqrt{\log x}\right).$$

Proof. Since α is not a Liouville number, we have for all $q \geq 1$,

$$\|\alpha q\| > C(\alpha) q^{-K}, \tag{9}$$

for some $K \geq 1$. Put

$$L = \exp\left(\sqrt{\log x}\right).$$

By the Erdös-Turàn theorem (see Theorem 5.5 of [9]), we have

$$D(x) \ll \frac{x}{L} + \sum_{\ell=1}^{L} \frac{1}{\ell} S_\ell\,, \tag{10}$$

where

$$S_\ell = \left| \sum_{n \leq x} e(\ell \alpha P(n)) \right|.$$

For each ℓ, let q_ℓ be the largest convergent denominator in the continued fraction expansion of $\alpha\ell$ not exceeding L^{4K}. Then $\|\ell q_\ell \alpha\| < L^{-4K}$, and so, by (9), $\ell q_\ell \gg L^4$; thus, $q \gg L^3$. We therefore have $L^3 \ll q_\ell < L^{4K}$. An easy modification of Theorem 7 then shows that

$$S_\ell \ll x(\log x)^4 L^{-3/2},$$

and therefore,

$$\sum_{\ell=1}^{L} \frac{1}{\ell} S_\ell \ll x(\log x)^5 L^{-3/2} \ll XL^{-1}.$$

The theorem now follows from (10). □

References

[1] W. Banks, A. Harcharras and I. E. Shparlinski, ‘Smooth values of shifted primes in arithmetic progressions’, *Michigan Math. J.*, **52** (2004), 603–618.

[2] W. Banks and I. E. Shparlinski, ‘Congruences and exponential sums with the Euler function’, *High Primes and Misdemeanours: Lectures in Honour of the 60th Birthday of Hugh Cowie Williams*, Fields Institute Communications, vol.41, Amer. Math. Soc., 2004, 49–60.

[3] W. Banks and I. E. Shparlinski, ‘Congruences and rational exponential sums with the Euler function’, *Rocky Mountain J. Math.*, (to appear).

[4] E. R. Canfield, P. Erdős and C. Pomerance, ‘On a problem of Oppenheim concerning “Factorisatio Numerorum”’, *J. Number Theory*, **17** (1983), 1–28.

[5] R. Crandall and C. Pomerance, *Prime numbers: A Computational perspective*, Springer-Verlag, Berlin, 2001.

[6] H. Davenport, *Multiplicative number theory*, 2nd ed., Springer-Verlag, New York 1980.

[7] G. H. Hardy and E. M. Wright, *An Introduction to the theory of numbers*, Fifth Edition, The Clarendon Press, Oxford University Press, New York, 1979.

[8] H. Halberstam and H.-E. Richert *Sieve methds*, Academic Press, London, UK, 1974.

[9] G. Harman, *Metric number theory*, Clarendon Press, Oxford, 1998.

[10] A. Hildebrand and G. Tenenbaum, ‘Integers without large prime factors’, *J. de Théorie des Nombres de Bordeaux*, **5** (1993), 411–484.

[11] A. Ivić, ‘On sums involving reciprocals of the largest prime factor of an integer, II’, *Acta Arith.*, **71** (1995), 229–251.

[12] H. Montgomery, ‘Primes in arithmetic progressions’, *Mich. Math. J.*, **17** (1970), 33-39.

[13] S.-M. Oon, ‘Pseudorandom properties of prime factors’, *Periodica Math. Hungarica*, **49** (2004), 45–63.

[14] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, 1995.

[15] I. M. Vinogradov, *The method of trigonometric sums in the theory of numbers*, translated, revised and annotated by A. Davenport and K.F.Roth, Interscience, New York, 1954.