Michigan Math. J. 53 (2005)

Distributional Properties of the Largest Prime Factor

WILLIAM D. BANKS, GLYN HARMAN,
& IGOR E. SHPARLINSKI

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 $q \ge 1$ and a real number z, we define $\mathbf{e}_q(z) = \mathbf{e}(z/q)$, where $\mathbf{e}(z) = \exp(2\pi i z)$ as usual.

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

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

For the case of q fixed, this question has been previously considered by Ivić [11]. However, the approach in [11] apparently does not 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 Oon [13] has studied the distribution of $P(n)$ over the congruence classes of a fixed modulus q in the case of n itself belonging 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 this estimate is no longer correct (even by an order of magnitude) for $q \ge \exp\bigl(3\sqrt{\log x \log\log x}\,\bigr)$.

In Section 4 we study the function

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

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 $\log q \le \log^{1/3} x$. Here, $\pi(x) = \#\{p \le x\}$. We expect that the matching lower bound $\varpi(x; q, a) \gg \pi(x)/\varphi(q)$ also holds for such q, or perhaps even

Received August 2, 2004. Revision received February 17, 2005.

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 $\varrho(x;q,a)$, we expect that the behavior of $\varpi(x;q,a)$ changes for larger values of q. Unfortunately, the scarcity of results about smooth shifted primes is an obstacle to proving this.

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

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

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] asserting 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 used in [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. Recall that $\mu(1) = 1$ and $\mu(m) = 0$ if $m \ge 2$ is not squarefree, and that $\mu(m) = (-1)^k$ 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 $P(n) \le y$. 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 [4, Thm. 3.1]; 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 [14, Chap. III.5, Thm. 1]).

LEMMA 2. *Let $u = (\log x)/(\log y)$, where $x \geq y > 0$. If $u \geq 1$, 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 and by $\mathcal{P}[w, x]$ the set of primes p such that $w \leq p \leq x$; for simplicity, we write $\mathcal{P}[x]$ for $\mathcal{P}[0, x]$. If the parameters $x \geq y > 0$ are fixed within a discussion, we also put $\mathcal{P}_m = \mathcal{P}[L_m, x/m]$ for all $m \geq 1$, where $L_m = \max\{y, P(m)\}$.

LEMMA 3. *Let $x \geq y > 0$. For any two functions $h(k)$ and $f(k)$ satisfying $\max\{|h(k)|, |f(k)|\} \leq 1$ for all positive integers k,*

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

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

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

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

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

which, together with (1), finishes the proof. □

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

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

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

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

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

We also need the *Bombieri–Vinogradov theorem*. See [6, Chap. 28], where the form of this theorem is slightly different from that of the following statement, which can be derived by partial summation.

LEMMA 5. *For every fixed number $A > 0$ there is a constant $B > 0$ such that, for all $x \geq 2$, 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

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

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

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

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. Let b be an integer coprime to q. We start with the following identity:

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

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

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

It follows that

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

where $\omega(q)$ is the number of distinct prime divisors of q. Here we have used the result of Landau that

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

(see e.g. [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

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

Observing that

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

and that, for all $q\le x$,

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

we obtain

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

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

LEMMA 7. *Uniformly for* $\exp(\log^{1/5}x)\le y\le x^{1/2}$ *and* $q\le\exp(\log^{6/7}y)$, *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

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

Using [1, Thm. 1] and 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

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

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

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

Consequently,

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

Since

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

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

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

and the result follows. □

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

LEMMA 8. *Let $\alpha \in \mathbb{R}$ be fixed, and suppose there are integers a, q with $q \ge 1$ and* $\gcd(a,q) = 1$ *and*

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

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

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

Finally, we also need the following "major arc" bound, which can be deduced via partial summation from the bound in [6, Chap. 26, p. 147].

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

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

In particular,

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

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

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

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

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

$$y = \exp(q^{\Delta}/2) \quad \text{and} \quad u = \frac{\log x}{\log y} = 2q^{-\Delta} \log x.$$

By Lemmas 2 and 3, we have

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

For any m with $mL_m \leq x$,

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

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. Hence, for all positive integers m with $x/m \geq L_m \geq y$, since $q \leq 2 \log^{1/\Delta} y$ it follows that the estimate

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

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

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

where

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

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

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

Therefore,

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

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

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

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 $M < x\exp(\log^{-2/3} x)$ and $M \ge x\exp(\log^{-2/3} x)$ separately. Thus, it follows that

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

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

It is clear that Theorem 1 is nontrivial only 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 because 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 $\log(x/a)/\log a = v + o(v)$, Lemma 1 allows us to derive that

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

and the result follows. □

In view of the lower bound given by 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 define

$$y = \exp\bigl(3\sqrt{\log x}\,\bigr) \quad\text{and}\quad u = \frac{\log x}{\log y} = \frac{\sqrt{\log x}}{3}.$$

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

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

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

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

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

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

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

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

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

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

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

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 e.g. [7, Cor. 2.4.1]; note that $q < x/m$ by our choice of q), and for such m we obtain

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

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

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

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

$$T_1 = \frac{x}{\log^2 x} \sum_{\substack{m\leq x^{2/3}\\ \gcd(am+1,q)=1}} \frac{1}{\varphi(m)},$$

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

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

$$T_2 \ll x\sum_{j=M}^{L} \frac{1}{e^j\log^2(x/e^jq)} \sum_{\substack{e^{j-1}\leq m<e^j\\ P(m)\leq x/e^{j-1}\\ \gcd(am+1,q)=1}} \frac{m}{\varphi(m)}$$

$$\ll \frac{\varphi(q)}{q}\,x\sum_{j=M}^{L}\frac{1}{\log^2(x/e^jq)}\exp\left(-\frac{\log(e^j)}{2\log(x/e^{j-1})}\right).$$

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

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

$$\ll \frac{\varphi(q)}{q}\frac{x}{\log^2 x}\sum_{j=M}^{L}\frac{\log^2 x}{\log^2(x/e^{j-1})}\exp\left(-\frac{\log x}{2\log(x/e^{j-1})}\right)$$

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

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*

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

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

Proof. Without loss of generality, we can also assume that $q \ge \log^8 x$ because otherwise the bound is trivial. 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

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

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, for all positive integers $m \le x/y$ we have

$$\sum_{p \in \mathcal{P}_m} \mathbf{e}_q(ap) \ll \frac{x}{m}(q^{-1/2} + x^{-1/5}m^{1/5} + q^{1/2}x^{-1/2}m^{1/2})\log^3 x$$

$$\ll \frac{x}{m}(q^{-1/2} + y^{-1/5} + q^{1/2}y^{-1/2})\log^3 x.$$

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 remarked previously, 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*:

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

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

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

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 e.g. [8, Thm. 272]). If $q \geq \log^{1/\Delta} x$ then the bound of the theorem is trivial, while for $q < \log^{1/\Delta} x$ we have

$$qx^{-q^{-\Delta/2}} \leq x^{-q^{-\Delta}} \quad \text{and} \quad q\exp(-c_1\log^{1/3} x) \ll \exp\bigl(-\tfrac{1}{2}c_1\log^{1/3} x\bigr).$$

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 given by Theorem 6 takes the form

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

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

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

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 then obtain a bound for $S_\alpha(x)$ that implies the sequence $\{\alpha P(n) : n \geq 1\}$ is uniformly distributed modulo 1 when α is irrational.

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

$$g = \max\bigl\{q_j : q_j < \exp\bigl(\sqrt{\log x}\,\bigr)\bigr\}.$$

Then

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

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

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

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

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

Hence

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

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

$$O(x\log^{-8} x) = O(xg^{-1/3}).$$

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 yields the upper bound

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

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; hence the parameter g in Theorem 7 tends to infinity with x.

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

$$D(x) = \sup_{0 \leq \gamma \leq 1} |N_\gamma(x) - \gamma x|,$$

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

$$N_\gamma(x) = \#\{n \leq x : \langle \vartheta_n \rangle \leq \gamma\}.$$

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 $\{\alpha P(n) : n \geq 1\}$ is uniformly distributed modulo* 1.

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

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

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

REMARK 4. Unfortunately, there is no hope of obtaining an explicit discrepancy bound in Corollary 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

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

When 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 $\{\alpha P(n) : n \leq x\}$. Then, provided that α is not a Liouville number, we have*

$$D(x) \ll x \exp\bigl(-\sqrt{\log x}\,\bigr).$$

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

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

for some $K \ge 1$. Put

$$L = \exp\bigl(\sqrt{\log x}\,\bigr).$$

By the Erdős–Turàn theorem (see [9, Thm. 5.5]),

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

where

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

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), $\ell q_\ell \gg L^4$; hence $q_\ell \gg L^3$. We thus 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,* Fields Inst. Commun., 41, pp. 49–60, Amer. Math. Soc., Providence, RI, 2004.

[3] ———, *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.

[5] H. Davenport, *Multiplicative number theory,* 2nd ed., Grad. Texts in Math., 74, Springer-Verlag, New York, 1980.

[7] H. Halberstam and H.-E. Richert, *Sieve methods,* London Math. Soc. Monogr. (N.S.), 4, Academic Press, London, 1974.

[8] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers,* 5th ed., Oxford Univ. Press, New York, 1979.

[9] G. Harman, *Metric number theory,* London Math. Soc. Monogr. (N.S.), 18, Oxford Univ. Press, New York, 1998.

[10] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors,* J. Théor. Nombres 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,* Michigan Math. J. 17 (1970), 33–39.
[13] S.-M. Oon, *Pseudorandom properties of prime factors,* Period. Math. Hungar. 49 (2004), 107–118.
[14] G. Tenenbaum, *Introduction to analytic and probabilistic number theory,* Cambridge Stud. Adv. Math., 46, Cambridge Univ. 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), Wiley, New York, 1954.

W. D. Banks
Department of Mathematics
University of Missouri
Columbia, MO 65211

bbanks@math.missouri.edu

G. Harman
Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX
United Kingdom

g.harman@rhul.ac.uk

I. E. Shparlinski
Department of Computing
Macquarie University
Sydney, NSW 2109
Australia

igor@ics.mq.edu.au