Mathematical Research Letters **11**, 853–868 (2004)

ALMOST ALL PALINDROMES ARE COMPOSITE

William D. Banks, Derrick N. Hart, and Mayumi Sakata

Never odd or even...

Abstract. We study the distribution of palindromic numbers (with respect to a fixed base $g \geq 2$) over certain congruence classes, and we derive a nontrivial upper bound for the number of prime palindromes $n \leq x$ as $x \to \infty$. Our results show that almost all palindromes in a given base are composite.

1. Introduction

Fix once and for all an integer $g \geq 2$, and consider the *base g representation* of an arbitrary natural number $n \in \mathbb{N}$:

$$n = \sum_{k=0}^{L-1} a_k(n) g^k.$$

Here $a_k(n) \in \{0, 1, \dots, g-1\}$ for each $k = 0, 1, \dots, L-1$, and we assume that the leading digit $a_{L-1}(n)$ is nonzero. The integer n is said to be a *palindrome* if its digits satisfy the symmetry condition:

$$a_k(n) = a_{L-1-k}(n), \qquad k = 0, 1, \dots, L-1.$$

Let $\mathcal{P} \subset \mathbb{N}$ denote the set of palindromes (in base g), and for every positive real number x, let

$$\mathcal{P}(x) = \{n \leq x \,|\, n \in \mathcal{P}\}.$$

In this paper, we study the distribution of palindromes in congruence classes. Using the Weil bound for mixed Kloosterman sums, we bound exponential sums over the set $\mathcal{P}_L$ of palindromes with precisely L digits and use this result to show that the set $\mathcal{P}(x)$ becomes uniformly distributed (as $x \to \infty$) over the congruence classes modulo p, where $p > g$ is any prime number for which the multiplicative order $\mathrm{ord}_p(g)$ of g in the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is at least $3p^{1/2}$; see Corollary 4.4 for a precise statement. We remark that, thanks to the work of Pappalardi [4], almost all primes p satisfy the stronger condition $\mathrm{ord}_p(g) \geq p^{1/2} \exp((\log p)^c)$ where c is any constant less than $(1 - \log 2)/2$; see also [2, 3].

Received April 24, 2003.
2000 *Mathematics Subject Classification.* 11A63, 11L07, 11N69.

Using a variation of these techniques, we also show that the set $\mathcal{P}(x)$ becomes uniformly distributed (as $x \to \infty$) over the congruence classes modulo q, where $q \ge 2$ is *any* integer coprime to $g(g^2-1)$; see Corollary 4.5. This latter result, though weaker than that obtained for primes p satisfying the condition $\mathrm{ord}_p(g) \ge 3p^{1/2}$, allows us to deduce the main result of this paper: *almost all palindromes in a given base are composite*. More precisely, in Theorem 5.1, we show that

$$\#\big\{n \in \mathcal{P}(x) \,|\, n \text{ is prime}\big\} = O\left(\#\mathcal{P}(x)\,\frac{\log\log\log x}{\log\log x}\right), \qquad x \to \infty,$$

where the implied constant depends only on the base g. This result appears to be the first of its kind in the literature.

2. Preliminary Estimates

For any integer $q \ge 2$, let $e_q(x)$ denote the exponential function $\exp(2\pi i x/q)$, which is defined for all $x \in \mathbb{R}$. For any integer c that is relatively prime to q, let $\overline{c}$ denote an arbitrary multiplicative inverse for c modulo q; that is, $c\overline{c} \equiv 1 \pmod q$. Finally, let $d(q)$ be the number of positive integral divisors of q, and let $\mathrm{ord}_q(g)$ be the smallest integer $t \ge 1$ such that $g^t \equiv 1 \pmod q$.

Lemma 2.1. *For any prime p with $\gcd(p,g) = 1$ and all $a, b \in \mathbb{Z}$, we have*

$$\left|\sum_{k=1}^{\mathrm{ord}_p(g)} e_p(ag^k + b\overline{g}^{\,k})\right| \le 2p^{1/2}\gcd(a,b,p)^{1/2}.$$

Proof. Consider the mixed Kloosterman sum

$$K_\chi(a,b;p) = \sum_{c=1}^{p-1} \chi(c)\, e_p(ac + b\overline{c}),$$

where χ is a Dirichlet character modulo p. From the work of Weil [6], it follows that the bound

$$\big|K_\chi(a,b;p)\big| \le 2p^{1/2}\gcd(a,b,p)^{1/2}$$

holds for all such sums. Averaging over all Dirichlet characters χ modulo p for which $\chi(g) = 1$, it follows that

$$\frac{\mathrm{ord}_p(g)}{\varphi(p)} \sum_\chi K_\chi(a,b;p) = \sum_{\substack{1 \le c \le p-1 \\ c \equiv g^k \pmod p,\ \exists k}} e_p(ac + b\overline{c}) = \sum_{k=1}^{\mathrm{ord}_p(g)} e_p(ag^k + b\overline{g}^{\,k}).$$

The result follows. □

Lemma 2.2. *The following bound holds for all $q \geq 2$, $k \geq 2$ and $h \in \mathbb{Z}$ provided that $q \nmid h$:*

$$\left|\sum_{a=0}^{k-1} e_q(ha)\right| \leq k \, \exp\left(-\frac{4\gcd(h,q)^2}{q^2}\right).$$

Proof. Let us write

$$s(q,k,h) = \left|\sum_{a=0}^{k-1} e_q(ha)\right|.$$

If $d = \gcd(h,q)$, then $s(q,k,h) = s(q/d,k,h/d)$, hence it suffices to prove the assertion for the special case where $\gcd(h,q) = 1$, which we now assume.

Without loss of generality, we may also suppose that $k \leq q$. Indeed, if $k \geq q+1$, then we can express $k = mq + r$ with $0 \leq r \leq q-1$ and simply observe that

$$s(q,k,h) = s(q,r,h) \leq r \leq q-1 \leq (q+1)\exp(-4/q^2) \leq k \, \exp(-4/q^2).$$

If $\gcd(h,q) = 1$ and $2 \leq k \leq q$, we have

$$\begin{aligned} s(q,k,h)^2 = \sum_{a,b=0}^{k-1} e_q\left(h(a-b)\right) &= k + \sum_{\substack{a,b=0 \\ a \neq b}}^{k-1} \cos\left(\frac{2\pi h(a-b)}{q}\right) \\ &\leq k + k(k-1)\cos(2\pi/q); \end{aligned}$$

therefore,

$$\frac{s(q,k,h)^2}{k^2} \leq \frac{1}{k} + \left(1 - \frac{1}{k}\right)\cos(2\pi/q) \leq \frac{1}{2}\left(1 + \cos(2\pi/q)\right).$$

Using the fact that $1 + \cos x \leq 2\exp(-x^2/4)$ for $0 \leq x \leq \pi$, we obtain the desired result. □

3. Exponential Sums over Palindromes

For every $L \geq 1$, let $\mathcal{P}_L$ denote the set of palindromes (in base g) with precisely L digits; that is,

$$\mathcal{P}_L = \{n \in \mathcal{P} \mid g^{L-1} \leq n < g^L\}.$$

Lemma 3.1. *Let $p > g$ be a prime number such that $\mathrm{ord}_p(g) > 2p^{1/2}$. Then for every $c \in \mathbb{Z}$ with $\gcd(c,p) = 1$, the exponential sum*

$$S_L(c) = \sum_{n \in \mathcal{P}_L} e_p(cn)$$

satisfies the bound

$$\left|S_L(c)\right| \leq {}^{\#}\mathcal{P}_L \cdot \Theta^{(L - 2\mathrm{ord}_p(g) - 1)/4},$$

where

$$\Theta = \frac{1}{g} + \frac{2(g-1)p^{1/2}}{g \operatorname{ord}_p(g)} < 1.$$

Proof. Since

$$\begin{aligned} S_{2L}(c) &= \sum_{a_0=1}^{g-1} \sum_{a_1=0}^{g-1} \dots \sum_{a_{L-1}=0}^{g-1} e_p\left(\sum_{k=0}^{L-1} ca_k \left(g^k + g^{2L-1-k}\right)\right) \\ &= \sum_{a_0=1}^{g-1} e_p\left(ca_0\left(1+g^{2L-1}\right)\right) \prod_{k=1}^{L-1} \sum_{a_k=0}^{g-1} e_p\left(ca_k\left(g^k + g^{2L-1-k}\right)\right) \end{aligned}$$

and

$$\begin{aligned} S_{2L+1}(c) &= \sum_{a_0=1}^{g-1} \sum_{a_1=0}^{g-1} \dots \sum_{a_L=0}^{g-1} e_p\left(ca_L g^L + \sum_{k=0}^{L-1} ca_k \left(g^k + g^{2L-k}\right)\right) \\ &= \sum_{a_0=1}^{g-1} e_p\left(ca_0(1+g^{2L})\right) \sum_{a_L=0}^{g-1} e_p(ca_L g^L) \prod_{k=1}^{L-1} \sum_{a_k=0}^{g-1} e_p\left(ca_k\left(g^k + g^{2L-k}\right)\right), \end{aligned}$$

it follows that

$$\left|S_{2L+\delta}(c)\right| \le (g-1)g^{\delta} \prod_{k=1}^{L-1} \left|\sum_{a=0}^{g-1} e_p\left(ca\left(g^k + g^{2L+\delta-1-k}\right)\right)\right|$$

for all $L \ge 1$ and $\delta = 0$ or 1.

Put $N = \operatorname{ord}_p(g)$, and write $L - 1 = Nm + \ell$, where $m = \lfloor (L-1)/N \rfloor$ and $0 \le \ell < N$. Then, using the arithmetic-geometric mean inequality, we derive that

$$\begin{aligned} \left|S_{2L+\delta}(c)\right|^2 &\le (g-1)^2 g^{2\ell+2\delta} \prod_{k=1}^{Nm} \left|\sum_{a=0}^{g-1} e_p\left(ca\left(g^k + g^{2L+\delta-1-k}\right)\right)\right|^2 \\ &\le (g-1)^2 g^{2\ell+2\delta} \left(\frac{1}{Nm} \sum_{k=1}^{Nm} \left|\sum_{a=0}^{g-1} e_p\left(ca\left(g^k + g^{2L+\delta-1-k}\right)\right)\right|^2\right)^{Nm} \\ &= (g-1)^2 g^{2\ell+2\delta} \left(\frac{T}{Nm}\right)^{Nm}, \end{aligned}$$

where

$$\begin{aligned} T &= \sum_{k=1}^{Nm} \sum_{a,b=0}^{g-1} e_p\left(c(a-b)\left(g^k + g^{2L+\delta-1-k}\right)\right) \\ &= gNm + m \sum_{\substack{a,b=0 \\ a \ne b}}^{g-1} \sum_{k=1}^{N} e_p\left(c(a-b)\left(g^k + g^{2L+\delta-1-k}\right)\right). \end{aligned}$$

Since $p > g$, it follows that $\gcd(a-b,p) = 1$ whenever $a \neq b$. Using Lemma 2.1, we therefore obtain that

$$\left|T\right| \leq gNm + (g-1)gm \cdot 2p^{1/2}.$$

Consequently,

$$\begin{aligned}\left|S_{2L+\delta}(c)\right|^2 &\leq (g-1)^2 g^{2\ell+2\delta} \left(\frac{gNm + 2(g-1)gmp^{1/2}}{Nm}\right)^{Nm} \\ &= (g-1)^2 g^{2Nm+2\ell+2\delta} \left(\frac{N + 2(g-1)p^{1/2}}{gN}\right)^{Nm} \\ &= (g-1)^2 g^{2L+2\delta-2} \Theta^{L-\ell-1}.\end{aligned}$$

Since ${}^{\#}\mathcal{P}_{2L+\delta} = (g-1)g^{L+\delta-1}$, the result follows. □

Lemma 3.2. *Let $q \geq 2$ be an integer such that $\gcd\left(q, g(g^2-1)\right) = 1$. Then for every $c \in \mathbb{Z}$ such that $q \nmid c$, the exponential sum*

$$S_L(c) = \sum_{n \in \mathcal{P}_L} e_q(cn)$$

satisfies the bound

$$\left|S_L(c)\right| \leq {}^{\#}\mathcal{P}_L \cdot \exp\left(-\frac{(L-5)\gcd(c,q)^2}{q^2}\right).$$

Proof. As in the proof of Lemma 3.1, we have

$$\left|S_{2L+\delta}(c)\right| \leq (g-1)g^{\delta} \prod_{k=1}^{L-1} \left|\sum_{a=0}^{g-1} e_q\left(ca\left(g^k + g^{2L+\delta-1-k}\right)\right)\right|$$

for all $L \geq 1$ and $\delta = 0$ or 1. Let

$$\begin{aligned}B &= \{1 \leq k \leq L-1 \,|\, q \text{ divides } c(g^k + g^{2L+\delta-1-k})\}, \\ G &= \{1 \leq k \leq L-1 \,|\, q \text{ does not divide } c(g^k + g^{2L+\delta-1-k})\}.\end{aligned}$$

Using Lemma 2.2 to estimate individual terms in the preceding product when $k \in G$, and using the trivial estimate when $k \in B$, we obtain that

$$\begin{aligned}\left|S_{2L+\delta}(c)\right| &\leq (g-1)g^{(\delta+{}^{\#}G+{}^{\#}B)} \exp\left(-\frac{4\gcd(c,q)^2\,{}^{\#}G}{q^2}\right) \\ &= {}^{\#}\mathcal{P}_{2L+\delta} \cdot \exp\left(-\frac{4\gcd(c,q)^2\,{}^{\#}G}{q^2}\right).\end{aligned}$$

Now let $f = q/\gcd(c,q)$. Since q does not divide c, we have $f \geq 2$, and the stated condition on q implies that $\mathrm{ord}_f(g^2) \geq 2$. Thus, if k and ℓ both lie in B, then

$$(g^2)^k \equiv -g^{2L+\delta-1} \equiv (g^2)^{\ell} \pmod{f}, \qquad k \equiv \ell \pmod{\mathrm{ord}_f(g^2)}.$$

We therefore see that

$$\begin{aligned} {}^{\#}B &\le 1 + \lfloor (L-2)/2 \rfloor = \lfloor L/2 \rfloor, \\ {}^{\#}G &\ge L - 1 - \lfloor L/2 \rfloor \ge L/2 - 1 \ge (2L + \delta - 5)/4, \end{aligned}$$

and the result follows. □

4. Distribution of Palindromes

Proposition 4.1. *Let $p > g$ be a prime number such that $\operatorname{ord}_p(g) \ge 3p^{1/2}$. Then for every $L \ge 10p - 5$, the following estimate holds for all $a \in \mathbb{Z}$:*

$$\left| {}^{\#}\big\{ n \in \mathcal{P}_L \,|\, n \equiv a \pmod p \big\} - \frac{{}^{\#}\mathcal{P}_L}{p} \right| < \frac{{}^{\#}\mathcal{P}_L}{p}\,(0.99)^L.$$

Proof. Using the relation

$$\frac{1}{p} \sum_{c=0}^{p-1} e_p(cm) = \begin{cases} 1 & \text{if } m \equiv 0 \pmod p, \\ 0 & \text{otherwise,} \end{cases}$$

it follows that

$$\begin{aligned} {}^{\#}\big\{ n \in \mathcal{P}_L \,|\, n \equiv a \pmod p \big\} &= \sum_{n \in \mathcal{P}_L} \frac{1}{p} \sum_{c=0}^{p-1} e_p\big(c(n-a)\big) \\ &= \frac{1}{p} \sum_{c=0}^{p-1} e_p(-ca) \sum_{n \in \mathcal{P}_L} e_p(cn) \\ &= \frac{{}^{\#}\mathcal{P}_L}{p} + \frac{1}{p} \sum_{c=1}^{p-1} e_p(-ca) S_L(c), \end{aligned}$$

where $S_L(c)$ is the exponential sum considered in Lemma 3.1. Therefore

$$\begin{aligned} \left| {}^{\#}\big\{ n \in \mathcal{P}_L \,|\, n \equiv a \pmod p \big\} - \frac{{}^{\#}\mathcal{P}_L}{p} \right| &\le \frac{1}{p} \sum_{c=1}^{p-1} \big| S_L(c) \big| \\ &\le \frac{{}^{\#}\mathcal{P}_L}{p} \sum_{c=1}^{p-1} \Theta^{(L - 2\operatorname{ord}_p(g) - 1)/4}, \end{aligned}$$

where

$$\Theta = \frac{1}{g} + \frac{2(g-1)p^{1/2}}{g \operatorname{ord}_p(g)} \le \frac{1}{g} + \frac{2(g-1)}{3g} = \frac{2g+1}{3g} \le \frac{5}{6},$$

since $g \ge 2$. Also,

$$(L - 2\operatorname{ord}_p(g) - 1)/4 \ge (L - 2p + 1)/4 \ge L/5, \qquad L \ge 10p - 5.$$

Consequently,

$$\left|{}^{\#}\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod p\} - \frac{{}^{\#}\mathcal{P}_L}{p}\right| \le \frac{{}^{\#}\mathcal{P}_L}{p}\,(p-1)\left(\frac{5}{6}\right)^{L/5}.$$

Finally, remarking that the condition $\operatorname{ord}_p(g) \ge 3p^{1/2}$ implies that $p \ge 11$, we have

$$(p-1)\left(\frac{5}{6}\right)^{L/5} < (0.99)^L, \qquad L \ge 10p - 5.$$

This completes the proof. □

Proposition 4.2. *Let $q \ge 2$ be an integer such that $\gcd\big(q, g(g^2-1)\big) = 1$. Then for every $L \ge 10 + 2q^2 \log q$, the following estimate holds for all $a \in \mathbb{Z}$:*

$$\left|{}^{\#}\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod q\} - \frac{{}^{\#}\mathcal{P}_L}{q}\right| < \frac{{}^{\#}\mathcal{P}_L}{q}\,\exp\left(-\frac{L}{2q^2}\right).$$

Proof. Using the relation

$$\frac{1}{q}\sum_{c=0}^{q-1} e_q(cm) = \begin{cases} 1 & \text{if } m \equiv 0 \pmod q, \\ 0 & \text{otherwise,} \end{cases}$$

it follows that

$$\begin{aligned} {}^{\#}\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod q\} &= \sum_{n \in \mathcal{P}_L} \frac{1}{q}\sum_{c=0}^{q-1} e_q\left(c(n-a)\right) \\ &= \frac{1}{q}\sum_{c=0}^{q-1} e_q(-ca) \sum_{n \in \mathcal{P}_L} e_q(cn) \\ &= \frac{{}^{\#}\mathcal{P}_L}{q} + \frac{1}{q}\sum_{c=1}^{q-1} e_q(-ca) S_L(c), \end{aligned}$$

where $S_L(c)$ is the exponential sum considered in Lemma 3.2. If $1 \le c \le q-1$, then $q \nmid c$, hence by Lemma 3.2 we derive the estimate:

$$\begin{aligned} \left|S_L(c)\right| &\le \frac{{}^{\#}\mathcal{P}_L}{q}\,\exp\left(\log q - \frac{(L-5)\gcd(c,q)^2}{q^2}\right) \\ &\le \frac{{}^{\#}\mathcal{P}_L}{q}\,\exp\left(\log q - \frac{L-5}{q^2}\right) \le \frac{{}^{\#}\mathcal{P}_L}{q}\,\exp\left(-\frac{L}{2q^2}\right), \end{aligned}$$

the last inequality following from the stated condition on L. The result follows immediately. □

Theorem 4.3. *Let $q \ge 2$ be a fixed integer, and suppose that there exist constants $A \ge 1$ and $\sqrt{2/3} \le \xi < 1$, depending only on q, such that*

$$\left|{}^{\#}\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod q\} - \frac{{}^{\#}\mathcal{P}_L}{q}\right| \le {}^{\#}\mathcal{P}_L \cdot A\,\xi^L$$

for all $L \ge 1$ and $a \in \mathbb{Z}$. Then for some constant $B \ge 1$ that depends only on g, the following estimate holds for all $x \ge 1$ and $a \in \mathbb{Z}$:

$$\left| {}^{\#}\big\{n \in \mathcal{P}(x) \,|\, n \equiv a \pmod q\big\} - \frac{{}^{\#}\mathcal{P}(x)}{q} \right| \le {}^{\#}\mathcal{P}(x) \cdot AB\, \xi^{(\log x)/(2 \log g)}.$$

Proof. We remark that the condition $\xi \ge \sqrt{2/3}$ guarantees that $g\xi^2$ is bounded below by an absolute constant greater than 1; since $g \ge 2$, we have

$$\frac{g-1}{g\xi^2 - 1} \le \frac{g-1}{\frac{2}{3}g - 1} \le 3.$$

For all $L \ge 1$, $x \ge y > 0$, and $a \in \mathbb{Z}$, let us denote

$$\begin{aligned} \mathcal{P}_a &= \{n \in \mathcal{P} \,|\, n \equiv a \pmod q\}, \\ \mathcal{P}_{a,L} &= \{n \in \mathcal{P}_a \,|\, g^{L-1} \le n < g^L\}, \\ \mathcal{P}_a(x) &= \{n \in \mathcal{P}_a \,|\, n \le x\}, \\ \mathcal{P}_a(y;x) &= \{n \in \mathcal{P}_a \,|\, y < n \le x\}. \end{aligned}$$

We also denote

$$\mathcal{P}(y;x) = \{n \in \mathcal{P} \,|\, y < n \le x\}.$$

In what follows, the implied constants in the symbol "O" may depend on g but are absolute otherwise. We recall that the notation $U = O(V)$ for positive functions U and V is equivalent to $U \le cV$ for some constant c.

Let $a \in \mathbb{Z}$ be fixed in what follows, and suppose that $g^{2M+\delta-1} \le x < g^{2M+\delta}$, where M is an integer and $\delta = 0$ or 1. We observe that

$$ {}^{\#}\mathcal{P}(x) = {}^{\#}\mathcal{P}(g^{2M+\delta-1}) + {}^{\#}\mathcal{P}(g^{2M+\delta-1}; x), \tag{1}$$

and that

$${}^{\#}\mathcal{P}_a(x) = {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}) + {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; x).$$

Our goal is to estimate

$$\begin{aligned} &\left| {}^{\#}\mathcal{P}_a(x) - \frac{{}^{\#}\mathcal{P}(x)}{q} \right| \\ &\le \left| {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}) - \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1})}{q} \right| + \left| {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; x) - \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1}; x)}{q} \right|. \end{aligned} \tag{2}$$

Since the integer $g^{2M+\delta-1}$ is *not* a palindrome (a fact that is only used to simplify our notation), we have by a straightforward calculation:

$$ {}^{\#}\mathcal{P}(g^{2M+\delta-1}) = g^M + g^{M+\delta-1} - 2. \tag{3}$$

On the other hand,

$$\begin{aligned}\#\mathcal{P}_a(g^{2M+\delta-1}) &= \sum_{L=1}^{2M+\delta-1} \#\mathcal{P}_{a,L} = \sum_{\ell=0}^{M-1} \#\mathcal{P}_{a,2\ell+1} + \sum_{\ell=1}^{M+\delta-1} \#\mathcal{P}_{a,2\ell} \\ &= \sum_{\ell=0}^{M-1} \left(\#\mathcal{P}_{a,2\ell+1} - \frac{\#\mathcal{P}_{2\ell+1}}{q} + \frac{\#\mathcal{P}_{2\ell+1}}{q} \right) + \sum_{\ell=1}^{M+\delta-1} \left(\#\mathcal{P}_{a,2\ell} - \frac{\#\mathcal{P}_{2\ell}}{q} + \frac{\#\mathcal{P}_{2\ell}}{q} \right) \\ &= \frac{\#\mathcal{P}(g^{2M+\delta-1})}{q} + \sum_{\ell=0}^{M-1} \left(\#\mathcal{P}_{a,2\ell+1} - \frac{\#\mathcal{P}_{2\ell+1}}{q} \right) + \sum_{\ell=1}^{M+\delta-1} \left(\#\mathcal{P}_{a,2\ell} - \frac{\#\mathcal{P}_{2\ell}}{q} \right).\end{aligned}$$

Using the hypothesis of the theorem, it therefore follows that

$$\left| \#\mathcal{P}_a(g^{2M+\delta-1}) - \frac{\#\mathcal{P}(g^{2M+\delta-1})}{q} \right| \le \sum_{\ell=0}^{M-1} \#\mathcal{P}_{2\ell+1} \cdot A\,\xi^{2\ell+1} + \sum_{\ell=1}^{M+\delta-1} \#\mathcal{P}_{2\ell} \cdot A\,\xi^{2\ell}.$$

Since

$$\begin{aligned}&\sum_{\ell=0}^{M-1} \#\mathcal{P}_{2\ell+1}\,\xi^{2\ell+1} + \sum_{\ell=1}^{M+\delta-1} \#\mathcal{P}_{2\ell}\,\xi^{2\ell} \\ &\qquad = \sum_{\ell=0}^{M-1} (g-1) g^{\ell} \xi^{2\ell+1} + \sum_{\ell=1}^{M+\delta-1} (g-1) g^{\ell-1} \xi^{2\ell} \\ &\qquad < \frac{g-1}{g\xi^2 - 1} \left(g^M \xi^{2M+1} + g^{M+\delta-1} \xi^{2M+2\delta} \right) = O\left(g^M \xi^{2M} \right),\end{aligned}$$

we see that

$$\left| \#\mathcal{P}_a(g^{2M+\delta-1}) - \frac{\#\mathcal{P}(g^{2M+\delta-1})}{q} \right| = O\left(A g^M \xi^{2M} \right). \tag{4}$$

We now turn to the more delicate estimation of $\#\mathcal{P}_a(g^{2M+\delta-1}; x)$. To this end, put $M = K + L$, where K and L are positive integers to be selected later. Examining the base g representation of an arbitrary palindrome n in $\mathcal{P}_{2M+\delta}$, we see that n may be expressed either in the form

$$n = n_1 + g^{K+\mu} n_2 + g^{K+2L+\delta} n_3,$$

or the form

$$n = n_1 + g^{K+2L+\delta} n_3,$$

where

$$1 \le n_1 < g^K, \qquad g^{K-1} \le n_3 < g^K, \qquad n_1 + g^K n_3 \in \mathcal{P}_{2K}, \tag{5}$$

and, in the former case, $n_2 \in \mathcal{P}_{2L+\delta-2\mu}$ for some $0 \le \mu \le L+\delta-1$. The integers n_1, n_2, n_3, μ are uniquely determined by n. We call n_3 the *K-signature* of n and write $s_K(n) = n_3$. The integer n_1 is uniquely determined by n_3 together with the first and third conditions of (5); we call n_1 the *K-complement* of n_3 and write $c_K(n_3) = n_1$.

Note that the number of palindromes $n \in \mathcal{P}_{2M+\delta}$ with a fixed K-signature $s_K(n) = n_3$ is precisely

$$1 + \sum_{\mu=0}^{L+\delta-1} {}^{\#}\mathcal{P}_{2L+\delta-2\mu} = 1 + \sum_{\mu=0}^{L+\delta-1} (g-1)g^{L+\delta-\mu-1} = g^{L+\delta}. \tag{6}$$

Now, given x in the range $g^{2M+\delta-1} \le x < g^{2M+\delta}$, let y be the palindrome in $\mathcal{P}_{2M+\delta}$ defined by

$$y = y_1 + g^K(g^{2L+\delta} - 1) + g^{K+2L+\delta} y_3,$$

where

$$y_3 = \begin{cases} \lfloor x/g^{K+2L+\delta} \rfloor + 1 & \text{if } g^{2M+\delta-1} \le x < g^{2M+\delta-1/2}, \\ \lfloor x/g^{K+2L+\delta} \rfloor - 1 & \text{if } g^{2M+\delta-1/2} \le x < g^{2M+\delta}, \end{cases}$$

and $y_1 = c_K(y_3)$. If x lies in the smaller range, then $x < y$, while $y < x$ if x lies in the larger range. In either case, we have

$$\bigl|{}^{\#}\mathcal{P}(g^{2M+\delta-1}; x) - {}^{\#}\mathcal{P}(g^{2M+\delta-1}; y)\bigr| = O(g^L) \tag{7}$$

and

$$\bigl|{}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; x) - {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; y)\bigr| = O(g^L),$$

since there are at most $O(1)$ distinct K-signatures for palindromes between x and y. Consequently,

$$\begin{aligned} &\left|{}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; x) - \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1}; x)}{q}\right| \\ &\qquad = \left|{}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; y) - \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1}; y)}{q}\right| + O(g^L). \end{aligned} \tag{8}$$

Now, if $n \in \mathcal{P}(g^{2M+\delta-1}; y)$, then its K-signature lies in the range

$$g^{K-1} \le s_K(n) \le y_3.$$

Thus,

$${}^{\#}\mathcal{P}(g^{2M+\delta-1}; y) = (y_3 - g^{K-1} + 1)g^{L+\delta}. \tag{9}$$

On the other hand, if $n \in \mathcal{P}_a(g^{2M+\delta-1}; y)$ with $s_K(n) = n_3$, then either

$$n = n_1 + g^{K+\mu} n_2 + g^{K+2L+\delta} n_3 \equiv a \pmod q$$

or

$$n = n_1 + g^{K+2L+\delta} n_3 \equiv a \pmod q,$$

depending on the form of n. In the latter case, there is at most one such palindrome n (for each fixed K-signature n_3), while in the former case, since

$$n_2 \equiv g^{-K-\mu}\bigl(a - c_K(n_3) - g^{K+2L+\delta} n_3\bigr) \pmod q,$$

the number of such palindromes n is ${}^{\#}\mathcal{P}_{b,2L+\delta-2\mu}$ for each $0 \le \mu \le L+\delta-1$, where

$$b = b(n_3, \mu) = g^{-K-\mu}(a - c_K(n_3) - g^{K+2L+\delta} n_3).$$

Hence, using (6), we derive that

$$\begin{aligned}
{}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; y) &= \sum_{n_3=g^{K-1}}^{y_3} \sum_{\mu=0}^{L+\delta-1} {}^{\#}\mathcal{P}_{b,2L+\delta-2\mu} + O(g^K) \\
&= \sum_{n_3=g^{K-1}}^{y_3} \left(\frac{1}{q} + \sum_{\mu=0}^{L+\delta-1} \frac{{}^{\#}\mathcal{P}_{2L+\delta-2\mu}}{q} \right) \\
&\qquad + \sum_{n_3=g^{K-1}}^{y_3} \sum_{\mu=0}^{L+\delta-1} \left({}^{\#}\mathcal{P}_{b,2L+\delta-2\mu} - \frac{{}^{\#}\mathcal{P}_{2L+\delta-2\mu}}{q} \right) + O(g^K) \\
&= \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1}; y)}{q} + \sum_{n_3=g^{K-1}}^{y_3} \sum_{\mu=0}^{L+\delta-1} \left({}^{\#}\mathcal{P}_{b,2L+\delta-2\mu} - \frac{{}^{\#}\mathcal{P}_{2L+\delta-2\mu}}{q} \right) + O(g^K).
\end{aligned}$$

Using the hypothesis of the theorem, it therefore follows that

$$\begin{aligned}
&\left| {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; y) - \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1}; y)}{q} \right| \\
&\qquad \le \sum_{n_3=g^{K-1}}^{y_3} \sum_{\mu=0}^{L+\delta-1} {}^{\#}\mathcal{P}_{2L+\delta-2\mu} \cdot A\,\xi^{2L+\delta-2\mu} + O(g^K) \\
&\qquad = \sum_{n_3=g^{K-1}}^{y_3} \sum_{\mu=0}^{L+\delta-1} (g-1) g^{L+\delta-\mu-1} \cdot A\,\xi^{2L+\delta-2\mu} + O(g^K) \\
&\qquad < A(y_3 - g^{K-1} + 1) \left(\frac{g-1}{g\xi^2 - 1}\, g^{L+\delta} \xi^{2L+\delta+2} \right) + O(g^K),
\end{aligned}$$

and consequently,

$$\left| {}^{\#}\mathcal{P}_a(g^{2M+\delta-1}; y) - \frac{{}^{\#}\mathcal{P}(g^{2M+\delta-1}; y)}{q} \right| = O\left(Ag^M \xi^{2L}\right) + O(g^K).$$

Using this estimate together with (2), (4) and (8), it follows that

$$\left| {}^{\#}\mathcal{P}_a(x) - \frac{{}^{\#}\mathcal{P}(x)}{q} \right| = O\left(Ag^M \xi^{2L} + g^L + g^K\right).$$

We now choose integers $K = M/2 + O(1)$ and $L = M/2 + O(1)$ such that $K + L = M$. Since $g\xi^2 > 1$ and $A \ge 1$, we have

$$\max\{g^K, g^L\} = O(g^{M/2}) = O\left(Ag^{M/2}(g\xi^2)^{M/2}\right) = O\left(Ag^M \xi^M\right),$$

therefore

$$\left| {}^{\#}\mathcal{P}_a(x) - \frac{{}^{\#}\mathcal{P}(x)}{q} \right| = O\left(Ag^M \xi^M\right).$$

To complete the proof, we need only observe that

$$\xi^M = O\left(\xi^{(\log x)/(2\log g)}\right)$$

for x in the range $g^{2M+\delta-1} \le x < g^{2M+\delta}$, and using (1), (3), (7) and (9) together with our choice of y_3, it follows that

$$\#\mathcal{P}(x) = g^M + \frac{x}{g^M} + O(g^{M/2});$$

thus $g^M = O\left(\#\mathcal{P}(x)\right)$. □

Using Theorem 4.3, we can now derive two immediate corollaries.

Corollary 4.4. *Let $p > g$ be a prime number such that $\operatorname{ord}_p(g) \ge 3p^{1/2}$. Then for some constant $C > 0$, depending only on g, the following estimate holds for all $x \ge 1$ and $a \in \mathbb{Z}$:*

$$\left|\#\left\{n \in \mathcal{P}(x) \,|\, n \equiv a \pmod p\right\} - \frac{\#\mathcal{P}(x)}{p}\right| \le \#\mathcal{P}(x) \cdot C\,(0.99)^{\frac{\log x}{2\log g} - 10p}.$$

Proof. Using the trivial estimate

$$\left|\#\left\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod p\right\} - \frac{\#\mathcal{P}_L}{p}\right| \le \#\mathcal{P}_L$$

for $1 \le L \le 10p - 6$, it follows from Proposition 4.1 that the estimate

$$\left|\#\left\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod p\right\} - \frac{\#\mathcal{P}_L}{p}\right| \le \#\mathcal{P}_L \cdot (0.99)^{L-10p+6}$$

holds for all $L \ge 1$ and $a \in \mathbb{Z}$. The result now follows immediately from Theorem 4.3. □

Corollary 4.5. *Let $q \ge 2$ be an integer such that $\gcd\left(q, g(g^2-1)\right) = 1$. Then for some constant $C > 0$, depending only on g, the following estimate holds for all $x \ge 1$ and $a \in \mathbb{Z}$:*

$$\left|\#\left\{n \in \mathcal{P}(x) \,|\, n \equiv a \pmod q\right\} - \frac{\#\mathcal{P}(x)}{q}\right| \le \#\mathcal{P}(x) \cdot C\,q \exp\left(-\frac{\log x}{4q^2 \log g}\right).$$

Proof. Using the trivial estimate

$$\left|\#\left\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod q\right\} - \frac{\#\mathcal{P}_L}{q}\right| \le \#\mathcal{P}_L$$

for $1 \le L < 10 + 2q^2 \log q$, it follows from Proposition 4.2 that the estimate

$$\left|\#\left\{n \in \mathcal{P}_L \,|\, n \equiv a \pmod q\right\} - \frac{\#\mathcal{P}_L}{q}\right| \le \#\mathcal{P}_L \exp\left(-\frac{(L - 10 - 2q^2\log q)}{2q^2}\right)$$

holds for all $L \ge 1$ and $a \in \mathbb{Z}$. The result now follows immediately from Theorem 4.3. □

5. Prime Palindromes

We now come to the main result of this paper.

Theorem 5.1. *As $x \to \infty$, we have*

$$\#\big\{n \in \mathcal{P}(x) \,|\, n \text{ is prime}\big\} = O\left(\#\mathcal{P}(x)\,\frac{\log\log\log x}{\log\log x}\right),$$

where the implied constant depends only on g.

Proof. As in the proof of Theorem 4.3, all implied constants in the symbol "O" may depend on g but are absolute otherwise.

Assuming that x is sufficiently large, let

$$h = \lfloor e \log\log\log x \rfloor, \qquad y = e^{-1}(\log x)^{1/4h} = \exp\left(\frac{\log\log x}{4e\log\log\log x}\right)^{1+o(1)}.$$

Let

$$Q = Q(y) = \prod_{g^3 < p \le y} p,$$

where the product runs over prime numbers. Note that $\gcd\big(Q, g(g^2-1)\big) = 1$. By Mertens' formula (see Theorem 11 in §I.1.6 of [5]), we have the estimate

$$\frac{\varphi(Q)}{Q} = \prod_{g^3 < p \le y}\left(1 - \frac{1}{p}\right) = O\big((\log y)^{-1}\big) = O\left(\frac{\log\log\log x}{\log\log x}\right), \tag{10}$$

where $\varphi(n)$ is the Euler function.

Now, if $n \in \mathcal{P}(x)$ is prime, either $\gcd(n, Q) = 1$ or n is a prime divisor of Q. We apply Brun's combinatorial sieve in the form given by Corollary 1.1 in §I.4.2 of [5]:

$$\#\big\{n \in \mathcal{P}(x) \,|\, n \text{ is prime}\big\} \le y + \sum_{\substack{q \,|\, Q \\ \omega(q) \le 2h}} \mu(q) A_q,$$

where $\mu(q)$ is the Möbius function, $\omega(q)$ is the number of distinct prime divisors of q, and

$$A_q = \#\{n \in \mathcal{P}(x) \,|\, n \equiv 0 \pmod q\}.$$

By Corollary 4.5, we see that

$$A_q = \frac{\#\mathcal{P}(x)}{q} + O\left(\#\mathcal{P}(x)\, q \exp\left(-\frac{\log x}{4q^2 \log g}\right)\right).$$

If $q \,|\, Q$ and $\omega(q) \le 2h$, then

$$q \le y^{2h} = \frac{(\log x)^{1/2}}{e^{2h}},$$

and since the number of such divisors q is bounded by y^{2h}, we have

$$\sum_{\substack{q \mid Q \\ \omega(q) \le 2h}} q \exp\left(-\frac{\log x}{4q^2 \log g}\right) \le \frac{\log x}{e^{4h}} \exp\left(-\frac{e^{4h}}{4 \log g}\right)$$

$$= \exp\left(\log\log x - 4h - \frac{e^{4h}}{4\log g}\right) = O\left(\frac{1}{\log x}\right),$$

since $h = \lfloor e \log\log\log x \rfloor$. Therefore,

$$\#\big\{n \in \mathcal{P}(x) \,|\, n \text{ is prime}\big\}$$

$$\le y + \#\mathcal{P}(x) \sum_{q \mid Q} \frac{\mu(q)}{q} + O\left(\#\mathcal{P}(x) \sum_{\substack{q \mid Q \\ \omega(q) > 2h}} \frac{1}{q} + O\left(\frac{\#\mathcal{P}(x)}{\log x}\right)\right).$$

Since $y = x^{o(1)}$ and $x^{1/2} = O\big(\#\mathcal{P}(x)\big)$, the first term in this estimate is negligible. Also, using (10), we have

$$\#\mathcal{P}(x) \sum_{q \mid Q} \frac{\mu(q)}{q} = \#\mathcal{P}(x) \prod_{g^3 < p \le y} \left(1 - \frac{1}{p}\right) = O\left(\#\mathcal{P}(x)\, \frac{\log\log\log x}{\log\log x}\right).$$

Finally, we have

$$\sum_{\substack{q \mid Q \\ \omega(q) > 2h}} \frac{1}{q} \le \sum_{\substack{q \mid Q \\ \omega(q) > 2h}} \frac{e^{\omega(q) - 2h}}{q} \le e^{-2h} \prod_{p \le y} (1 + e/p) \le \exp\left(-2h + e \sum_{p \le y} 1/p\right).$$

Observing that

$$\sum_{p \le y} \frac{1}{p} = (\log\log y)(1 + o(1)) = (\log\log\log x)(1 + o(1)),$$

by our choice of h it follows that

$$\#\mathcal{P}(x) \sum_{\substack{q \mid Q \\ \omega(q) > 2h}} \frac{1}{q} \le \#\mathcal{P}(x) \exp\big((\log\log\log x)(-e + o(1))\big) = O\left(\frac{\#\mathcal{P}(x)}{(\log\log x)^2}\right).$$

This completes the proof. □

6. Remarks and Open Problems

Using estimates from [1], it is possible to establish a version of Lemma 3.1 under the weaker assumption that $\mathrm{ord}_p(g) \gg \log p$; this yields analogues of Proposition 4.1 and Corollary 4.4, however the uniform constant 0.99 in those results must be replaced by a term like $\exp(-(\log\log p)^{-c})$ for some constant $c > 0$.

It seems natural to conjecture that the set of palindromes should behave as "random" integers, thus one might expect that the asymptotic relation

$$\#\big\{n \in \mathcal{P}(x) \,|\, n \text{ is prime}\big\} \sim C\,\frac{\#\mathcal{P}(x)}{\log x}$$

holds for some constant $C > 0$. While this question seems out of reach at the moment, it should be feasible to derive the upper bound

$$\#\big\{n \in \mathcal{P}(x) \,|\, n \text{ is prime}\big\} = O\left(\frac{\#\mathcal{P}(x)}{\log x}\right)$$

using more sophisticated sieving techniques coupled with better estimates for the distribution of palindromes in congruence classes. It is still an open problem to show the existence of infinitely many prime palindromes for any fixed base $g \geq 2$.

Acknowledgments

The authors would like to thank Florian Luca and Igor Shparlinski, whose valuable observations on the original manuscript led to significant improvements in our estimates. During the preparation of this paper, W. B. was supported in part by NSF grant DMS-0070628.

References

[1] T. Cochrane, C. Pinner and J. Rosenhouse, *Bounds on exponential sums and the polynomial Waring problem mod p*, J. London Math. Soc. (2) **67** (2003), 319–336.

[2] P. Erdős and R. Murty, *On the order of a* (mod p), Number theory (Ottawa, ON, 1996), 87–97, CRM Proc. Lecture Notes **19**, Amer. Math. Soc., Providence, RI, 1999.

[3] K.-H. Indlekofer and N. Timofeev, *Divisors of shifted primes*, Publ. Math. Debrecen **60** (2002), 307–345.

[4] F. Pappalardi, *On the order of finitely generated subgroups of* Q^* (mod p) *and divisors of* $p-1$, J. Number Theory **57** (1996), 207–222.

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

[6] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U.S.A. (1948), 204–207.

Dept. of Mathematics, University of Missouri, Columbia, MO 65211, USA

E-mail address: bbanks@math.missouri.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

E-mail address: hart@math.gatech.edu

Dept. of Mathematics, William Jewell College, Liberty, MO 64068, USA

E-mail address: sakatam@william.jewell.edu