SOME REMARKS ON GAUSS SUMS ASSOCIATED WITH KTH POWERS

H. L. Montgomery[1], R. C. Vaughan[2,3] and T. D. Wooley[2,4]

1. Introduction

Estimates for rational trigonometric sums are of great importance in analysing the local aspects of many additive problems. Indeed, bounds for the sums

$$S(q,a)=\sum_{x=1}^{q}\mathrm{e}(ax^k/q),$$

in which $\mathrm{e}(\alpha)$ denotes $\exp(2\pi i\alpha)$, play a fundamental rôle in the application of the Hardy-Littlewood method to Waring's problem (see [11]), and also in the analysis of the local solubility of systems of additive equations (see, for example, [2]). When $k\ge 2$ is an integer, and p is a prime number it is well known (see [5] or [11, Lemma 4.3]) that

$$|S(p,a)|\le((k,p-1)-1)p^{1/2}\quad(p\not|a).\tag{1}$$

When $p<((k,p-1)-1)^2$, the latter estimate is worse than trivial. Since

$$\sum_{x=1}^{p}\mathrm{e}(ax^k/p)=\sum_{y=1}^{p}\mathrm{e}(ay^{(k,p-1)}/p),$$

we may suppose without loss of generality that

$$p\equiv 1\pmod{k}.\tag{2}$$

Then it is plausible, at least when k is growing at a reasonably rapid rate with p, that

$$S(p,a)\ll_\varepsilon (kp)^{1/2+\varepsilon}\quad(p\not|a),\tag{3}$$

but in general this seems to be well beyond our capabilities at present. Nevertheless, it is possible in many circumstances to improve on the estimate (1). In §2 below we establish the following theorem.

Theorem 1. *Let k be an even positive natural number. Suppose that p is an odd prime number with $p\equiv 1 \pmod{k}$, $2k\not|p-1$ and $p\not|a$. Then*

$$|S(p,a)|\le 2^{-1/2}(k^2-2k+2)^{1/2}p^{1/2}.$$

We note that the theorem gives an improvement over the classical bound (1) precisely when -1 is not a kth power $\pmod{p}$. This is significant because the contrary case is very often easier to handle in applications. Furthermore, when k is even, the theorem improves on (1) for around half of the primes p with $p\equiv 1\pmod{k}$.

By way of illustration, we record here a consequence of our new estimate for the local solubility of additive equations.

[1] Supported by NSF grant DMS-9107605

[2] Research supported in part by the Isaac Newton Institute for Mathematical Sciences

[3] Supported by a SERC Senior Fellowship

[4] Supported by NSF grant DMS-9303505, an Alfred P. Sloan Research Fellowship and a Fellowship from the David and Lucile Packard Foundation

Typeset by $\mathcal{AMS}$-TEX

Corollary. *Let r, l, n be positive integers with $l > 1$ and $n > 2rl$. Suppose that p is a prime which satisfies the property that -1 is not an lth power $\pmod p$. Suppose also that*

$$p > \tfrac{1}{2} l^{2r+2} \left(1 + \frac{1}{(l-1)^2}\right).$$

Then the system of equations

$$a_{i1} x_1^l + \cdots + a_{in} x_n^l = 0 \quad (1 \le i \le r),$$

with coefficients $a_{ij} \in \mathbb{Z}$, has a non-trivial p-adic solution.

In the absence of the condition that -1 not be an lth power modulo p, [2, Theorem 1] provides the same conclusion with the condition $p > \frac{1}{2} l^{2r+2} \left(1 + (l-1)^{-2}\right)$ replaced by $p > l^{2r+2}$. Despite the modest scale of our improvement, it should nonetheless prove useful in computational approaches to local solubility problems exemplified in the work of Cook (see, for example, [1,3]). We shall not provide a proof of the corollary, since on replacing the classical estimate (1) by Theorem 1, we may apply precisely the same argument as was used in the proof of [2, Theorem 1]. (We note that forthcoming work of the third author supersedes [2, Theorem 1] when $r > 1$.)

It may be worth pointing out that a standard argument involving exponential sums, in combination with Theorem 1, provides an improvement on the bounds of Weil [12] on the number of solutions of additive equations over finite fields.

It is remarkable, in view of the many applications of the sums $S(p, a)$, that little serious attention has been given to the problem of providing a credible model for their behaviour. Indeed, heuristic arguments are frequently based on nothing more substantial than the expectation of "square root cancellation". In §§3 and 4 below we examine the possibility of stronger bounds than (1), and in particular provide a probabilistic model for the behaviour of the sums $S(p, a)$. Based on the latter deliberations, we feel justified in making the following conjecture, which we expect to be close to the truth.

Conjecture. *Let k be an integer exceeding 2, and suppose that p is a prime number with $p \equiv 1 \pmod k$ and $p \not| a$. Then*

$$|S(p, a)| \le \min \left\{ (k-1) p^{1/2} \ , \ (1 + \eta(p, k)) \left(2kp \log(kp)\right)^{1/2} \right\},$$

where $\eta(p, k) \to 0$ as $(p/k, k) \to (\infty, \infty)$.

We note that, in response to the encouragement of the referee, we have performed extensive computations which lend credibility to our conjecture. We provide a brief report on these computations in the final section.

Throughout, $\ll$ and $\gg$ will denote Vinogradov's well known notation, implicit constants being absolute, unless otherwise indicated. We use some well known properties of Gauss sums in §2. All of the relevant material will be found in Davenport [4].

2. Proof of Theorem 1

Let k be an integer with $k > 1$. Suppose that p is a prime number with $k | p - 1$, $2k \not| p - 1$ and $p \not| a$. Then necessarily k is even, say $k = 2d$. Let $\mathcal{A}$ denote the set of $k - 1$ non-principal characters modulo p of order k. Then

$$S(p, a) = \sum_{\chi \in \mathcal{A}} \overline{\chi}(a) \tau(\chi), \tag{4}$$

where $\tau(\chi)$ is the Gauss sum

$$\tau(\chi) = \sum_{x=1}^{p} \chi(x) \mathrm{e}(x/p).$$

We characterise the elements $\chi \in \mathcal{A}$ by taking a primitive root g modulo p, and defining $\chi = \chi_r$ by $\chi_r(g^s) = \mathrm{e}(rs/k)$ $(1 \le r \le k-1)$. Then, by (4),

$$S(p, a) = \chi_d(a) \tau(\chi_d) + \sum_{r=1}^{d-1} \left(\overline{\chi}_r(a) \tau(\chi_r) + \chi_r(a) \tau(\overline{\chi}_r) \right).$$

Also, for each r we have $\chi_r(-1) = \pm 1$, and further, $\tau(\overline{\chi}_r) = \chi_r(-1)\overline{\tau(\chi_r)}$. Thus

$$S(p,a) = \chi_d(a)\tau(\chi_d) + 2 \sum_{\substack{1\le r\le d-1\\ \chi_r(-1)=1}} \Re\left(\overline{\chi}_r(a)\tau(\chi_r)\right) + 2i \sum_{\substack{1\le s\le d-1\\ \chi_s(-1)=-1}} \Im\left(\overline{\chi}_s(a)\tau(\chi_s)\right). \tag{5}$$

Now observe that $\chi_r(-1) = -1$ if and only if $\mathrm{e}(r(p-1)/(2k)) = -1$. Then since, by assumption, $2k \not| p-1$, the latter holds precisely when r is odd. The number of such r with $1 \le r \le d-1$ is $\left[\frac{1}{2}d\right]$. Similarly, $\chi_r(-1) = 1$ precisely when r is even, and the number of such r with $1 \le r \le d-1$ is $\left[\frac{1}{2}(d-1)\right]$. As for the initial term on the right hand side of equation (5), we have

$$\tau(\chi_d) = \begin{cases} p^{1/2}, & \text{for } p \equiv 1 \pmod 4,\\ ip^{1/2}, & \text{for } p \equiv 3 \pmod 4.\end{cases}$$

Thus, since for each r we have $|\tau(\chi_r)| = p^{1/2}$, we deduce from (5) that

$$|S(p,a)| \le 2p^{1/2}\left(\left(\tfrac{1}{2} + \left[\tfrac{1}{2}(d-e)\right]\right)^2 + \left[\tfrac{1}{2}(d-f)\right]^2\right)^{1/2},$$

where (e,f) is $(1,0)$ or $(0,1)$ according to whether $p \equiv 1 \pmod 4$ or $p \equiv 3 \pmod 4$. In the former instance we have $4|k$, and in the latter we have $2|k$ but $4 \not| k$. Thus in either case,

$$|S(p,a)| \le 2p^{1/2}\left(\left(\tfrac{1}{2}(d-1)\right)^2 + \left(\tfrac{1}{2}d\right)^2\right)^{1/2} = \left(k^2 - 2k + 2\right)^{1/2}(p/2)^{1/2},$$

which completes the proof of Theorem 1.

We note that Theorem 1 could also be established by using an argument applied by Mitkin [7] to prove an essentially unrelated result. However, the above treatment has the benefit of making explicit the dependence on Gauss sums, and this we shall find useful in §3 below.

3. The possibility of stronger bounds: arithmetic perspective

Any improvement of the bounds (1) and Theorem 1 in the direction of (3) would have a significant impact in applications. Thus we feel obliged to comment on the extent to which such an improvement may actually exist. We consider two perspectives, arithmetic and probabilistic.

We begin by investigating the arithmetic perspective. Define the characters χ_r as in §2, and consider the decomposition (4). Suppose first that $2k|p-1$. When k is even we have $p \equiv 1 \pmod 4$, and hence $\tau(\chi_{k/2}) = p^{1/2}$. Further, for each r we have $|\tau(\chi_r)| = p^{1/2}$, so we may choose $\theta_r \in [0,1)$ so that $\tau(\chi_r) = p^{1/2}\mathrm{e}(\theta_r)$. Let ω be 0 or 1 according to whether k is odd or even. Then as in (4),

$$\frac{S(p,g^t)}{p^{1/2}} = \omega\mathrm{e}(t/2) + 2\sum_{1\le r<k/2} \cos\left(2\pi\left(\theta_r - rt/k\right)\right). \tag{6}$$

Meanwhile, when $2k \not| p-1$, from equation (5) we obtain

$$\frac{S(p,g^t)}{p^{1/2}} = \omega\mathrm{e}(t/2) + \sum_{\substack{1\le r<k/2\\ r\ \text{even}}} 2\cos\left(2\pi\left(\theta_r - rt/k\right)\right) + \sum_{\substack{1\le s<k/2\\ s\ \text{odd}}} 2i\sin\left(2\pi\left(\theta_s - st/k\right)\right), \tag{7}$$

where now ω is 1 or i according to whether $p \equiv 1 \pmod 4$ or $p \equiv 3 \pmod 4$.

The strength of the bounds achievable for $S(p,a)$ plainly depends on the distribution of the θ_r $(1 \le r < k/2)$. We distinguish three cases.

$k = 3$. In this case (6) reduces to

$$S(p,g^t) = 2p^{1/2}\cos\left(2\pi\left(\theta_1 - t/3\right)\right).$$

Heath-Brown and Patterson [6] have shown that the θ_1 are uniformly distributed modulo 1 as p ranges over the prime numbers $p \equiv 1 \pmod 3$. Then $S(p,g^t)$ is infinitely often arbitrarily close to each number in the interval $[-2p^{1/2}, 2p^{1/2}]$.

$k = 4$. In this case the equations (6) and (7) reduce to

$$S(p, g^t) = \begin{cases} \mathrm{e}(t/2)p^{1/2} + 2p^{1/2}\cos\left(2\pi\left(\theta_1 - t/4\right)\right), & \text{when } p \equiv 1 \pmod 8, \\ \mathrm{e}(t/2)p^{1/2} + 2ip^{1/2}\sin\left(2\pi\left(\theta_1 - t/4\right)\right), & \text{when } p \equiv 5 \pmod 8. \end{cases}$$

Work of Patterson [9] shows that the θ_1 are uniformly distributed modulo 1 as p ranges over the prime numbers $p \equiv 1 \pmod 4$. Furthermore, it seems likely that after some technical complications, Patterson's method can be modified to establish the same uniform distribution result with $p \equiv 1 \pmod 8$, or alternatively, with $p \equiv 5 \pmod 8$. Consequently, for primes in the former congruence class, $|S(p, a)|$ will be infinitely often arbitrarily close to each number in the interval $[0, 3p^{1/2}]$, and for primes in the latter congruence class, $|S(p, a)|$ will be infinitely often arbitrarily close to each number in the interval $[p^{1/2}, (5p)^{1/2}]$.

$k > 4$. In this case the situation is less clear. Patterson [9] has shown that for each r, the θ_r are uniformly distributed as p ranges over the primes $p \equiv 1 \pmod k$. A priori, however, it is not impossible that the θ_r may be mutually dependent in some manner for all such primes, and this could lead to considerable cancellation. One suspects that as p varies and k is kept fixed, the θ_r with $1 \le r < k/2$ come arbitrarily close to any given configuration. If so, by (6) and (7), the modulus of $S(p, a)$ is infinitely often arbitrarily close to $\Delta p^{1/2}$, where $\Delta = k - 1$ for the congruence class $p \equiv 1 \pmod{2k}$, and $\Delta = 2^{-1/2}(k^2 - 2k + 2)^{1/2}$ for the class $p \equiv k + 1 \pmod{2k}$. At present, unfortunately, it seems that even the distribution of a given non-trivial linear form in θ_r and θ_s, when $1 \le r < s < k/2$, is not known to be uniform.

The assertions, for large k, contained in the last paragraph are consequences of the hypothesis that the arguments of the Gauss sums are quasi-randomly distributed. Thus, in a strong sense, the above "arithmetic model" is merely a probabilistic model for the Fourier transforms of the $S(p, a)$.

4. The possibility of stronger bounds: a probabilistic model

Let k be an integer with $k > 1$, let p denote a prime number with $p \equiv 1 \pmod k$, and take g to be a primitive root modulo p. Then on writing

$$s = (p - 1)/k$$

we have

$$S(p, g^v) = 1 + k\sum_{u=1}^{s} \mathrm{e}(g^{uk+v}/p),$$

and

$$\sum_{v=1}^{k} |S(p, g^v)|^2 = k(k - 1)p.$$

Thus the $|S(p, a)|$ are typically about $(kp)^{1/2}$. This lends credibility to the estimate (3), a conjecture we shall examine through the use of a probabilistic model (see also [8] for related probabilistic considerations).

When k and s are not too small we expect the residues g^{uk+v} modulo p to be fairly randomly distributed on $[1, p-1] \cap \mathbb{Z}$ with respect to u and, also, independently with respect to v. Consequently we surmise that the terms $\mathrm{e}(g^{uk+v}/p)$ are somewhat randomly distributed on the unit circle with argument roughly uniformly distributed with respect to each parameter. Let ϕ_{uv} $(1 \le u \le s, 1 \le v \le k)$ be mutually independent random variables on $[-\frac{1}{2}, \frac{1}{2}]$, each with uniform distribution function. Then an appropriate probabilistic model for $S(p, g^v)$ is given by then random variable $\sigma_v(p)$, which we define by

$$\sigma_v(p) = 1 + k\sum_{u=1}^{s} \mathrm{e}(\phi_{uv}).$$

For the purposes of the analysis which follows we make the simplifying assumption that $2k | p - 1$. Then -1 is a kth power residue modulo p, so we may rewrite $S(p, g^v)$ in the form

$$S(p, g^v) = 1 + 2k\Re\left(\sum_{u=1}^{s/2} \mathrm{e}(g^{uk+v}/p)\right).$$

Thus in this case it is appropriate to model $S(p, g^v)$ by the random variable

$$\tau_v(p) = 1 + 2k \sum_{u=1}^{s/2} \cos(2\pi\phi_{uv}). \tag{8}$$

We note that it should be possible to treat $\sigma_v(p)$ in a manner similar to $\tau_v(p)$, with similar conclusions, by considering its real and imaginary parts separately. However, in view of the complications which arise in this process, we restrict our attention to $\tau_v(p)$.

Theorem 2. *Suppose that the random variables $\tau_v(p)$, given by (8), are independent as p varies over the set $\mathcal{P}$ of primes in the arithmetic progression $p \equiv 1 \pmod{2k}$. Suppose also that when $x \geq X$,*

$$\textit{card}\{p \in \mathcal{P} \ : \ x < p \leq 2x\} \asymp \frac{x}{\phi(2k)\log x}.$$

Further, suppose that $\varepsilon > 0$, that k is sufficiently large in terms of ε, and that

$$P \geq \max\left\{X \ , \ k(\log k)^{3+\varepsilon}\right\}. \tag{9}$$

Then

$$\max_{\substack{P<p\leq 2P \\ p\in\mathcal{P}}} \max_{1\leq v\leq k} \frac{|\tau_v(p)|}{(2kp\log p)^{1/2}} \geq 1 + \eta_-(k) \quad \textit{almost surely}, \tag{10}$$

where $\eta_-(k) \to 0$ as $k \to \infty$, and further

$$\sup_{\substack{p\geq P \\ p\in\mathcal{P}}} \max_{1\leq v\leq k} \frac{|\tau_v(p)|}{(2kp\log p)^{1/2}} \leq 1 + \eta_+(k) \quad \textit{almost surely}, \tag{11}$$

where $\eta_+(k) \to 0$ as $k \to \infty$.

Proof. In considering the behaviour of $\tau_v(p)$, we are naturally led to the Central Limit Theorem. However, the latter lacks sufficient uniformity to be of use to us in this instance, and we are therefore forced to apply the theory of large deviations. We begin with some observations concerning the random variables $\xi_{uv} = \cos(2\pi\phi_{uv})$. The probability density function, $f(\xi)$, of ξ_{uv}, is given by

$$f(\xi) = \begin{cases} \frac{1}{\pi}(1-\xi^2)^{-1/2}, & \text{when } |\xi| < 1, \\ 0, & \text{when } |\xi| \geq 1. \end{cases}$$

Thus for each u and v it follows that ξ_{uv} has mean 0 and variance $\frac{1}{2}$. Further, plainly, there exist positive constants b and c such that $\Pr(|\xi_{uv}| \geq x) \leq b\mathrm{e}^{-cx}$ for all $x \geq 0$. Thus the ξ_{uv} satisfy the Cramér condition, by the preamble to Petrov [10, VIII.2, Theorem 1] (see also [10, III.4, Lemma 5]).

We bound $|\tau_v(p) - 1|$ through estimates for $\alpha_v(p)$, which we define by

$$\alpha_v(p) = k^{-1}s^{-1/2}(\tau_v(p) - 1) = 2s^{-1/2}\sum_{j=1}^{s/2} \xi_{jv}. \tag{12}$$

Let $F_s(x)$ be the probability distribution function of $\alpha_v(p)$, and let $\Phi(x)$ be the normal distribution function corresponding to mean 0 and variance 1, that is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\mathrm{d}t. \tag{13}$$

Then since the Cramér condition is satisfied, it follows from [10, VIII.2, Theorem 1] that for each $x \geq 0$ with

$$x = o(s^{1/2}),$$

we have

$$\log\left(\frac{1 - F_s(x)}{1 - \Phi(x)}\right) = x^3(s/2)^{-1/2}\lambda(x(s/2)^{-1/2}) + O\left((x+1)s^{-1/2}\right), \tag{14}$$

and

$$\log\left(\frac{F_s(-x)}{\Phi(-x)}\right) = -x^3(s/2)^{-1/2}\lambda(-x(s/2)^{-1/2}) + O\left((x+1)s^{-1/2}\right), \tag{15}$$

were, $\lambda(t)$, the Cramér series, is a power series in t which converges for sufficiently small values of $|t|$. Thus there is a function $\Lambda_s(x) : \mathbb{R} \to \mathbb{R}_{>0}$ which satisfies, when $xs^{-1/2}$ is sufficiently small,

$$\Lambda_s(x) = \exp\left(x^3(s/2)^{-1/2}\lambda\left(x(s/2)^{-1/2}\right) + O\left((|x|+1)s^{-1/2}\right)\right), \tag{16}$$

and such that for each $\gamma > 0$ with $\gamma s^{-1/2}$ sufficiently small, we have

$$\Pr(\alpha_v(p) > \gamma) = 1 - F_s(\gamma) = \Lambda_s(\gamma)(1 - \Phi(\gamma))$$

and

$$\Pr(\alpha_v(p) \le -\gamma) = F_s(-\gamma) = \Lambda_s(-\gamma)(1 - \Phi(\gamma)).$$

Thus

$$\Pr(\alpha_v(p) \le \gamma) = \Phi(\gamma) + \operatorname{sgn}(\gamma)(1 - \Lambda_s(\gamma))(1 - \Phi(|\gamma|)),$$

and so for $\beta/(ks)$ sufficiently small we have

$$\Pr(\tau_v(p) \le \beta) = \Phi\left(\frac{\beta-1}{ks^{1/2}}\right) + \operatorname{sgn}(\beta-1)\left(1 - \Lambda_s\left(\frac{\beta-1}{ks^{1/2}}\right)\right)\left(1 - \Phi_s\left(\frac{|\beta-1|}{ks^{1/2}}\right)\right). \tag{17}$$

Therefore, when $\beta > 1$ we have

$$\Pr(|\tau_v(p)| > \beta) = \Phi\left(\frac{-\beta+1}{ks^{1/2}}\right)\Lambda_s\left(\frac{\beta-1}{ks^{1/2}}\right) + \Phi\left(\frac{-\beta-1}{ks^{1/2}}\right)\Lambda_s\left(\frac{-\beta-1}{ks^{1/2}}\right). \tag{18}$$

On hypothesis we have $k(\log k)^3p^{-1}$ small. Suppose that $\frac{1}{2} \le \tau \le 3$. Then $\tau(kp\log p)^{1/2}(ks)^{-1}$ is small and

$$\tau(kp\log p)^{1/2}k^{-1}s^{-1/2} = \tau(p(\log p)/(p-1))^{1/2}.$$

Thus we deduce from (13), (16) and (18) that

$$\Pr\left(|\tau_v(p)| > \tau(kp\log p)^{1/2}\right) = p^{-\tau^2/2 + O(\log\log p/\log p)}.$$

Let $\psi(x)$ denote a positive function of the positive variable x which decreases monotonically to 0 sufficiently slowly as x tends to ∞. Then for p large enough we have

$$\Pr\left(|\tau_v(p)| > ((1 \pm \psi(p))2kp\log p)^{1/2}\right) = p^{-1\mp(1+o(1))\psi(p)}. \tag{19}$$

We define the random variables $X_p^\pm$ by

$$X_p^\pm = \begin{cases} 1, & \text{when } |\tau_v(p)| \le ((1 \pm \psi(p))2kp\log p)^{1/2} \text{ for } 1 \le v \le k, \\ 0, & \text{otherwise.} \end{cases}$$

Then since the ϕ_{uv} are independent with respect to v, it follows from (19) that for each prime $p \in \mathcal{P}$, we have

$$\Pr\left(X_p^\pm = 1\right) = \left(1 - p^{-1\mp(1+o(1))\psi(p)}\right)^k. \tag{20}$$

Now define the random variable $Z_P^\pm$ by

$$Z_P^\pm = \begin{cases} 1, & \text{when } X_p^\pm = 1 \text{ for each } p \in \mathcal{P} \text{ with } P < p \le 2P, \\ 0, & \text{otherwise.} \end{cases}$$

Then since the $X_p^{\pm}$ are independent as p varies, it follows from (20) that

$$\Pr\left(Z_P^{\pm}=1\right)=\prod_{\substack{P<p\le 2P\\ p\in\mathcal{P}}}\left(1-p^{-1\mp(1+o(1))\psi(p)}\right)^k.$$

Suppose that P satisfies (9). Then on recalling our hypothesis on the distribution of the primes $p\in\mathcal{P}$, we deduce that

$$\Pr\left(Z_P^{\pm}=1\right)=\exp\left(-\frac{P^{\mp(1+o(1))\psi(P)}}{\log P}\frac{k}{\phi(2k)}\right)=\exp\left(-P^{\mp(1+o(1))\psi(P)}\right). \tag{21}$$

Then in particular,

$$\Pr\left(Z_P^{-}=1\right)=\exp\left(-P^{(1+o(1))\psi(P)}\right)=o(1),$$

and thus the event $Z_P^-=1$ almost surely does not occur. Consequently, when P satisfies (9), it is almost surely the case that for some $p\in\mathcal{P}$ with $P<p\le 2P$, and some v with $1\le v\le k$, we have

$$|\tau_v(p)|>\left((1-\psi(p))2kp\log p\right)^{1/2},$$

and thus we have established (10). Similarly, we deduce from (21) that

$$\Pr\left(Z_P^{+}=1\right)=\exp\left(-P^{-(1+o(1))\psi(P)}\right)=1+o(1),$$

and thus the event $Z_P^+=1$ almost surely occurs. Moreover,

$$\Pr(Z_{2^jP}^{+}=1\text{ for each }j\ge 0)=\prod_{j=0}^{\infty}\exp\left(-(2^jP)^{-(1+o(1))\psi(2^jP)}\right)=1+o(1),$$

and hence when P satisfies (9), it is almost surely the case that

$$|\tau_v(p)|>\left((1+\psi(p))2k(p-1)\log p\right)^{1/2}$$

holds for no $p\in\mathcal{P}$ with $p>P$, and no v with $1\le v\le k$. This establishes (11), and completes the proof of the theorem.

Thus far we have not varied k in the analysis. One might expect that the terms $\mathrm{e}(g^{uk+v}/p)$ demonstrate quasi-random behaviour with respect to k. Such an assumption may be examined through an analysis similar to that above, provided that one makes additional assumptions concerning independence of this behaviour. An appropriate treatment would replace (19) with the conclusion

$$\Pr\left(\left|\tau_v^{(k)}(p)\right|>\left((1\pm\psi(p))2kp\log(kp)\right)^{1/2}\right)=(pk)^{-1\mp(1+o(1))\psi(p)},$$

where we have now made the dependence on k explicit. With the definitions of $X_p^{\pm}$ and $Z_P^{\pm}$ modified accordingly, we now define the random variable $W_{P,D}^{\pm}$ by

$$W_{P,D}^{\pm}=\begin{cases}1, & \text{when } Z_P^{\pm}=1 \text{ for each } k \text{ with } D<k\le 2D,\\ 0, & \text{otherwise.}\end{cases}$$

Then

$$\begin{aligned}\Pr\left(W_{P,D}^{\pm}=1\right)&=\prod_{D<k\le 2D}\prod_{\substack{P<p\le 2P\\ p\in\mathcal{P}_k}}\left(1-(pk)^{-1\mp(1+o(1))\psi(p)}\right)^k\\ &=\exp\left(-(PD)^{\mp(1+o(1))\psi(P)}\right).\end{aligned}$$

In this way one is able to draw conclusions similar to those in Theorem 2. In particular, when p and k are sufficiently large, almost surely one has

$$\tau_v^{(k)}(p) < ((1+o(1))2kp\log(kp))^{1/2}.$$

Such a conclusion, we argue, provides some evidence in support of the conjecture made in the introduction.

The most fundamental assumptions necessary for the probabilistic model to give estimates close to the truth concern the uniform distribution, and independence, of the kth powers modulo primes in the congruence class $p \equiv 1 \pmod{2k}$. Unfortunately, while these assumptions are entirely plausible, and widely held to be true, there is as yet no strong evidence in their favour. On the other hand, the hypothesis in the statement of Theorem 2 concerning the distribution of primes is entirely benign. Presumably, one could comfortably take $X = k(\log k)^A$, for a sufficiently large A, in the hypothesis. As far as unconditional results go, it is an immediate consequence of the Siegel-Walfisz theorem that one may take $X = \exp(k^\varepsilon)$ for any $\varepsilon > 0$, and indeed the sharpest modern forms of the Bombieri-Vinogradov theorem would permit one to take $X = k^2$ for almost all k.

We note that the above probabilistic argument can also be applied to the decompositions (6) and (7), these depending on the arguments of the Gauss sums. If we assume that the θ_r defined in §3 are rather uniformly distributed with respect to r, and also, independently, with respect to p, then we can set up a probabilistic model for the $S(p, g^v)$ in much the same way as above. A similar analysis will yield the same conclusions. This observation adds weight to our conjecture, but is also significant in another respect. It is clear that the probabilistic model inherent in Theorem 2 loses credibility when $p > \mathrm{e}^{k/2}$, for then the model's conclusion (10) contradicts the bound (1). The explanation for this is simply that when p is very large, arithmetic effects generate some non-random structure amongst the kth powers. In these circumstances the Gauss sum based model remains effective, and is more appropriate. On the other hand, when $p < k^2$, a probabilistic model based on Gauss sums is flawed in that it permits the possibility that the model for $S(p, a)$ has modulus exceeding p. This occurs simply because when p is very small, arithmetic structure is genuinely dominated by quasi-random behaviour. Thus, in a strong sense, the two probabilistic models described apply to different, but overlapping, regimes.

Exponential sums of the more general form

$$\sum_{x=1}^{p} \mathrm{e}\left((a_1x + \cdots + a_kx^k)/p\right)$$

can also be examined through a model similar to that above. For such sums, the probabilistic model based on Gauss sums, alluded to in the previous paragraph, must be replaced by an analogous one arising from the well known work of Weil [13] on the Riemann Hypothesis for finite fields. However, the problem of determining to what extent the arguments of the summands are uniformly distributed and independent, as we vary the polynomial, is in general one of great difficulty.

5. Numerical Evidence

We have made quite extensive calculations of $S(p, g^v)$ for k up to $30{,}000$ and prime numbers $p \equiv 1$ modulo $2k$ up to $12{,}000{,}000$. For example, in Fig. 1 below we give a histogram of the distribution of $S(p, g^v)$ when $k = 25{,}431$ and $p = 10{,}324{,}987$ as v varies in $[1, k] \cap \mathbb{Z}$. More precisely, let $m = \min_v S(p, g^v)$ and $M = \max_v S(p, g^v)$ and put $\Delta = M - m$. Then we cover $[m, M]$ with about $\sqrt{k}$ intervals of length $\Delta/\sqrt{k}$. The histogram counts the number of $S(p, g^v)$ which occur in each particular interval. By (17), the expected number of $S(p, g^v)$ in a particular interval (β_1, β_2) is approximately $k\left(\Phi(\beta_2/k\sqrt{s}) - \Phi(\beta_1/k\sqrt{s})\right)$. Since $(\beta_2 - \beta_1)/k\sqrt{s} = \Delta/k\sqrt{p-1}$ is small, this is approximately

$$\frac{\Delta}{\sqrt{2\pi(p-1)}}\mathrm{e}^{-\frac{x^2}{2k(p-1)}}$$

for each $x \in (\beta_1, \beta_2)$. The smooth curve in Fig. 1 is precisely this function. This is fairly typical of the situation when p/k is large, and we believe that this provides good evidence in favour of our model for the general distribution of $S(p, g^v)$.

Our main interest, of course, lies in large values. There is an attendant difficulty which arises when we attempt to match the numerical evidence with the theory. This is that the errors which occur are

Fig. 1

likely to be inflated when we raise our estimates to the k-th power, as in (20). To this end it is useful to have further information regarding the Cramér series $\lambda(t)$. Following the proof of Theorem 2 of Petrov [10, VIII.2, Theorem 2], we find that $\lambda(t)$ is as follows. Let $I(z)$ denote the modified Bessel function

$$I_0(z) = \int_0^\pi \exp(z\cos\theta)\mathrm{d}\theta = \sum_{k=0}^\infty \frac{(z/2)^{2k}}{(k!)^2},$$

and for brevity drop the suffix 0, and put

$$L(z) = \log I(z).$$

Then $L'(z) = I'(z)/I(z)$ can be shown to be strictly increasing on $\mathbb{R}$ with infimum and supremum -1 and +1 respectively. Thus for each real number t with $|t| < \sqrt{2}$ there is a unique number $z = z(t)$ such that

$$t = \sqrt{2}L'(z).$$

The Cramér series $\lambda(t)$ is then defined by

$$\lambda(t) = t^{-3}\left(\frac{t^2}{2} - \frac{tz}{\sqrt{2}} + L(z)\right).$$

It is readily shown that this expression has a removable singularity at the origin, that $\lambda \in \mathcal{C}^\infty(-\sqrt{2}, \sqrt{2})$, and that $\lambda(0) = 0$. Moreover $\lambda(-t) = -\lambda(t)$. Below we give a table of values of $\lambda(t)$.

t	$z(t)$	$\lambda(t)$	t	$z(t)$	$\lambda(t)$
0.00	0.00	0.00	0.75	1.25783394	-0.05610566
0.10	0.14177639	- 0.00626742	0.80	1.38238761	-0.06163766
0.15	0.21333658	- 0.00943407	0.85	1.52000921	-0.06769428
0.20	0.28571919	- 0.01264090	0.90	1.67446014	-0.07439928
0.25	0.35922593	- 0.01590246	0.95	1.85115236	-0.08191987
0.30	0.43418401	- 0.01923434	1.00	2.05821540	-0.09048946
0.35	0.51095504	- 0.02265353	1.05	2.30851797	-0.10044718
0.40	0.58994601	- 0.02617881	1.10	2.62379143	-0.11231064
0.45	0.67162315	- 0.02983132	1.15	3.04381253	-0.12692051
0.50	0.75652993	- 0.03363516	1.20	3.64909559	-0.14575403
0.55	0.84531094	- 0.03761821	1.25	4.62466707	-0.17168403
0.60	0.93874435	- 0.04181326	1.30	6.48100952	-0.21113180
0.65	1.03778692	- 0.04625941	1.35	11.28118688	-0.28263997
0.70	1.14363815	- 0.05100406	1.40	50.00260469	-0.50652303

Table 1.

The theory as it stands requires in order to establish (14) and (15) that the expression $x/\sqrt{s}$ be small. Indeed, in our application we have x^4/s small, and then the expressions on the right of (14) and (15) are small. However, in performing calculations it is not at all clear what "small" should be. Fortunately, Table 1 indicates that for much of its range the function $\lambda(t)$ is relatively small. We thus take the liberty of replacing the right side of (17) by

$$\Phi\left(\frac{\beta}{ks^{1/2}}\right).$$

Thus for a given k and a given pair of numbers X and Y with $X < Y$ we consider the prime numbers p with $X < p \le Y$ and lying in the residue class 1 modulo $2k$. Let Π denote the number of such primes and put $\delta = 2/\sqrt{\Pi}$. For each prime p and each v we have

$$\Pr\left(\frac{|\tau_v(p)|}{\sqrt{k(p-1)}} \le \beta\right) = 2\Phi(\beta) - 1.$$

Thus assuming independence we have

$$\Pr(M_p \le \beta) = (2\Phi(\beta) - 1)^k ,$$

where $M_p = \max\{|\tau_1(p)|, \dots , |\tau_k(p)|\}/\sqrt{k(p-1)}$. In the histogram in Fig. 2 below we take $k = 2,987$, $X = 10^6$, $Y = 10,346,969$ and for $m = 1, 2, 3, \dots$ plot against the interval $[m-1, m]$ the number of primes p for which $\max\{|S(p, g^v)|\}/\sqrt{k(p-1)}$ lies in the interval $[(m-1)\delta, m\delta)$. The bare vertical line above each interval is the corresponding expected value

$$\Pi\left((2\Phi(m\delta) - 1)^k - (2\Phi((m-1)\delta) - 1)^k\right).$$

This is typical for reasonably large values of k and X/k. The fit is sufficiently close to give some confidence to the model when it is applied to large values. In particular the spread is almost invariably within the predicted limits.

References

1. O. D. Atkinson and R. J. Cook, *Pairs of additive congruences to a large prime modulus*, J. Austral. Math. Soc. **46A** (1989), 438–455.
2. O. D. Atkinson, J. Brüdern and R. J. Cook, *Simultaneous additive congruences to a large prime modulus*, Mathematika **39** (1992), 1–9.
3. R. J. Cook, *Pairs of additive congruences: cubic congruences*, Mathematika **32** (1986), 286–300.
4. H. Davenport, *Multiplicative number theory, 2nd Ed.*, Springer-Verlag, Berlin, 1980.
5. G. H. Hardy and J. E. Littlewood, *A new solution of Waring's problem*, Quart. J. Math. Oxford (2) **48** (1919), 272–293.
6. D. R. Heath-Brown and S. J. Patterson, *The distribution of Kummer sums at prime arguments*, J. Reine Angew. Math. **310** (1979), 111–130.
7. D. A. Mitkin, *Estimates of rational trigonometric sums of a special form*, Dokl. Akad. Nauk SSSR **224** (1975), 760–763.
8. R. W. K. Odoni, *The statistics of Weil's trigonometric sums*, Proc. Cambridge Philos. Soc. **74** (1973), 467–471.
9. S. J. Patterson, *The distribution of general Gauss sums and similar arithmetic functions at prime arguments*, Proc. Lond. Math. Soc. (3) **54** (1987), 193–215.
10. V. V. Petrov, *Sums of independent random variables*, Springer-Verlag, Berlin, 1975.
11. R. C. Vaughan, *The Hardy-Littlewood Method*, Cambridge University Press, Cambridge, 1981.
12. A. Weil, *Numbers of solutions of equations in finite fields*, Bull. Amer. Math. Soc. **55** (1949), 497–508.
13. A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Actualités Sci. Ind., No. 41 = Publ. Inst. Math. Univ. Strasbourg **7** (1945), Hermann et Cie, Paris 1948.

HLM: Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1003, U.S.A.
E-mail address: hlm@math.lsa.umich.edu

RCV: Department of Mathematics, Imperial College of Science and Technology, Queen's Gate, London SW7 2AZ, England
E-mail address: rvaughan@ma.ic.ac.uk

TDW: Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1003, U.S.A.
E-mail address: wooley@math.lsa.umich.edu

Fig. 2