On the Distribution of Multiplicative Translates of Sets of Residues (mod p)

J. Håstad

Royal Institute of Technology
Stockholm, Sweden

J. C. Lagarias

A. M. Odlyzko

AT&T Bell Laboratories
Murray Hill, NJ 07974

(July 29, 1992)

1. Introduction

Let p be a prime and view residues (mod p) as members of the set $\{0, 1, \ldots, p-1\}$. Let R be a set of r distinct nonzero residues (mod p). Suppose that the random variable a is drawn uniformly from $(\mathbb{Z}/p\mathbb{Z})^* = \mathbb{Z}/p\mathbb{Z} - \{0\}$, and denote the normalized expected size of the minimal residue in the multiplicative translate aR (mod p) by

$$M^*(R) := \frac{1}{p} E[\min(aR)] \ . \tag{1.1}$$

Our object is to obtain upper and lower bounds for $M^*(R)$ which depend only on $r = |R|$.

This problem arises in two contexts. The first concerns the analysis of a simple randomization scheme to select an element of a set $\tilde{R}$ of distinct integers in $[0, p-1]$ whose size $|\tilde{R}|$ is not specified in advance. Draw a and b independently from the uniform distributions on $(\mathbb{Z}/p\mathbb{Z})^*$ and $\mathbb{Z}/p\mathbb{Z}$, respectively. The set $a\tilde{R}$ divides the circle $\mathbb{R}/p\mathbb{Z}$ into $|\tilde{R}|$ intervals. The randomization procedure is to select that element $x \in \tilde{R}$ such that ax is the leftmost element of the interval in which b falls, i.e. ax is the largest element in $a\tilde{R}$ (mod p) less than or equal to b, unless there is no such integer, in which case ax is the largest element in $a\tilde{R}$ (mod p). The induced distribution on $\tilde{R}$ need not be uniform. How non-uniform can this induced distribution be? Since translating $\tilde{R}$ by an additive constant does not change the distribution, we may reduce to the case where $\tilde{R}$ contains 0, so that $\tilde{R} := R < \{0\}$, and investigate the probability that the

randomly selected element $x \in \tilde{R}$ equals 0. Thus we are interested in bounding the quantity

$$M(\tilde{R}) := \text{Prob}\{b = \min[aR + b] : a, b \in Z/pZ,\ a \neq 0\} . \tag{1.2}$$

This problem easily reduces to bounding quantities of the form (1.1), because one has

$$M(\tilde{R}) = M^*(-R) . \tag{1.3}$$

To see this, note that for fixed a and variable b, the element b is the smallest nonnegative residue of $a\tilde{R} + b$ exactly for $0 \le b \le p - \max[aR] = \min[-aR]$. Averaging over all a then gives (1.3).

The second context is the analysis of a particular pseudorandom number generator. Given a small number of absolutely random bits, called a *seed,* the pseudorandom number generation problem is to use these bits in a deterministic manner to produce a much larger number of ''random-looking'' bits. Here ''random-looking'' means that the bits appear well-distributed with respect to certain statistical measures. We call such a set of bits *pseudorandom bits* with respect to these measures, cf. Lagarias (1990). This problem can also be reversed: one can consider a particular deterministic mapping with random input and obtain bounds on the distribution of its output with respect to various statistical measures.

We consider a problem of this latter sort, concerning the generation procedure which when given a seed (a,b), consisting of elements a and b drawn independently from the uniform distributions on $(Z/pZ)^*$ and Z/pZ, respectively, together with a deterministically constructed set R, produces the set

$$aR + b \pmod{p}$$

as a set of $|R|$pseudorandom numbers. Here the seed (a, b) contains about $2\log_2 p$ random bits, while the output has about $r \log p$ bits, which can give an exponential expansion of the number of bits if $r = p^{\beta}$ with $\beta > 0$. The elements of $aR + b$ are pseudorandom only in a relatively weak

sense, but they do possess some nice distribution properties which are useful in applications. Indeed, it is well-known that if $\mathsf{R} = \{x_1, x_2\}$ consists of exactly two distinct elements, then the random variables $ax_1 + b$ and $ax_2 + b$ are independent and identically distributed if a and b are chosen independently from $\mathbb{Z}/p\mathbb{Z}$. Consequently the elements of $a\mathsf{R} + b$ are pairwise independent when regarded as random variables. This pairwise independence property has been exploited by Luby (1985) in constructing a simple parallel algorithm for the maximal independent set problem. See also Alon, Babai and Itai (1986), §6, for a history and applications of this idea to derandomize parallel algorithms. A related construction of k-wise independent variables is due to Joffe (1974), see also Zuckerman (1990).

Relevant statistical measures of $a\mathsf{R} + b$ in these applications concern the distribution of the lengths of the r intervals into which $a\mathsf{R} + b$ cuts the circle $\mathbb{R}/p\mathbb{Z}$, for $a \in (\mathbb{Z}/p\mathbb{Z})^*$, $b \in \mathbb{Z}/p\mathbb{Z}$. We consider here the *mean square spacing measure*

$$\mathrm{mss}[a\mathsf{R} + b] := \sum_{i=1}^{r} \lambda_i^2 ,$$

where λ_i are the lengths of these intervals, and the associated statistical measure

$$S(\mathsf{R}) := E[\mathrm{mss}[a\mathsf{R} + b]] . \tag{1.4}$$

The quantity $S(\mathsf{R})$ has a simple relation to various quantities $M^*(\mathsf{R}')$. Since the measure $S(\mathsf{R})$ is translation-invariant, one has

$$S(\mathsf{R}) = E[\mathrm{mss}[a\mathsf{R}]] . \tag{1.5}$$

Now set $\mathsf{R}_b := \mathsf{R} + b \pmod p$. One has the identities

$$\sum_{b=0}^{p-1} \min[\mathsf{R} + b] = \sum_{i=1}^{r} \frac{(\lambda_i - 1)\lambda_i}{2} = \frac{1}{2}\mathrm{mss}(\mathsf{R}) - \frac{1}{2}p$$

and

$$\sum_{b=0}^{p-1} \min[a\mathsf{R} + b] = \sum_{b=0}^{p-1} \min[a(\mathsf{R} + b)] ,$$

since $a \neq 0$. These yield

$$\begin{aligned} S(\mathsf{R}) &= \frac{1}{p-1} \sum_{a=0}^{p-1} \mathrm{mss}[a\mathsf{R}] \\ &= \frac{1}{p-1} \sum_{a=1}^{p-1} \left\{ 2 \sum_{b=0}^{p} \min[a(\mathsf{R} + b)] + p \right\} \\ &= 2 \sum_{b=0}^{p} pM^*(\mathsf{R}_b) + 2p . \end{aligned} \tag{1.6}$$

Thus $S(\mathsf{R})$ is determined by values of $M^*(\mathsf{R}_b)$ for various sets R_b having a fixed cardinality r. Consequently upper and lower bounds valid for all $M^*(\mathsf{R})$ of fixed cardinality r yield upper and lower bounds for all $S(\mathsf{R})$. It is possible that $S(\mathsf{R})$ satisfies stronger bounds than those inferred from $M^*(\mathsf{R})$, due to the averaging in (1.6). However if Conjecture 1.3 below is true, then little improvement is possible.

Now we describe our bounds for $M^*(\mathsf{R})$. For reference observe that the *expected size* of $M^*(\mathsf{R})$, averaged over all sets of cardinality r, is easily calculated to be

$$E[M^*(\mathsf{R}) : |\mathsf{R}| = r] = \frac{1}{r+1} , \tag{1.7}$$

because this average is exactly the expected size of the minimal element of a uniformly drawn r-tuple of $\{1, 2, \ldots, p-1\}$.

There is a simple lower bound for $M^*(\mathsf{R})$.

Theorem 1.1. *For all sets* R *(mod p) of cardinality r,*

$$M^*(\mathsf{R}) \geq \frac{1}{2r} - \frac{1}{pr} . \tag{1.8}$$

Proof. For any residue $x \in \{1, 2, \ldots, p-1\}$ there are exactly r values of a such that $x \in a\mathsf{R}$. Hence $\min[a\mathsf{R}] = x$ can occur at most r times, and $\min[a\mathsf{R}] = 0$ occurs once,

whence

$$M^*(\mathsf{R}) = \frac{r}{p}\left\lfloor \left(\sum_{j=1}^{\left\lceil \frac{p-1}{r} \right\rceil} j\right) + \left(\left\lceil \frac{p-1}{r} \right\rceil + 1\right)\left(p-1-r\left\lceil \frac{p-1}{r} \right\rceil\right)\right\rfloor$$

$$\geq \frac{r}{p}\,\frac{\frac{p-1}{r}\left(\frac{p-1}{r}+1\right)}{2}\,,$$

which gives (1.8). ■

This worst-case lower bound (1.8) for $M^*(\mathsf{R})$ gives something away, but in view of (1.7) it can be at most a multiplicative factor of 2, as $p \to \infty$. In fact, it is at most a smaller multiplicative factor, because in §3 we show that the set $\mathsf{J}_{2r} = \{\pm 1, \pm 2, \ldots, \pm r\}$ has

$$M^*(\mathsf{J}_{2r}) = c_r^* \left\lceil \frac{p}{2r} \right\rceil + O\left\lceil \frac{r^2}{p} \right\rceil, \tag{1.9}$$

for constants c_r^* satisfying

$$c_r^* = \frac{12 \log 2}{\pi^2} + O\left\lceil \frac{\log r}{r} \right\rceil$$

as $r \to \infty$. Hence the multiplicative factor is asymptotically at most $\frac{24 \log 2}{\pi^2} = 1.6855\ldots$ (Note that J_{2r} has $2r$ elements.)

The more interesting problem concerns worst-case upper bounds for $M^*(\mathsf{R})$. In §2, we establish the following bound.

Theorem 1.2. *For all primes p and for all sets* R *(mod p) of cardinality r,*

$$M^*(\mathsf{R}) \leq \frac{100}{r^{1/2}}\,. \tag{1.10}$$

The proof uses a second-moment method. The constant 100, as well as many of the other constants in §2, can be improved easily.

In §3 we show by example that the worst-case upper bound cannot be the same order of magnitude as (1.7). The set $\mathsf{I}_r = \{1, 2, \ldots, r\}$ has

$$M^*(\mathsf{I}_r) = c_r \frac{\log r}{r} + O\left[\frac{r^2}{p^2}\right] \tag{1.11}$$

where c_r are positive constants bounded away from zero, which satisfy

$$c_r = \frac{\pi^2}{24} + O\left[\frac{\log r}{r}\right]$$

as $r \to \infty$.

Another example is given by taking the set N_p of quadratic nonresidues (mod p), so that $r = (p-1)/2$. Then $\min[a\mathsf{N}_p]$ equals 1 if a is a quadratic nonresidue and otherwise it equals the minimal quadratic nonresidue α_p, so that

$$M^*[\mathsf{N}_p] = \frac{1}{2}(1 + \alpha_p) \ . \tag{1.12}$$

Graham and Ringrose (1990) show that α_p is infinitely often greater than $c^*(\log p)(\log\log\log p)$, for some constant $c^* > 0$, so that

$$M^*(\mathsf{N}_p) >> \left[\frac{r}{\log r \log\log\log r}\right]^{-1} \tag{1.13}$$

for such primes p. If the Generalized Riemann Hypothesis is true, then the $\log\log\log r$ factor in (1.13) can be strengthened to $\log\log r$.

What is the true order-of-magnitude of the worst-case bound for $M^*(\mathsf{R})$? For definiteness we propose:

Conjecture 1.3. *For each $\varepsilon > 0$ there is a constant $c(\varepsilon)$ such that for all primes p and all sets*

$\mathsf{R} \pmod p$,

$$M^*(\mathsf{R}) \le c(\varepsilon) r^{-1+\varepsilon} . \tag{1.14}$$

This conjecture is likely to be hard to settle affirmatively, in view of the quadratic nonresidue example N_p. Proving (1.14) for $\mathsf{R} = \mathsf{N}_p$ and $\varepsilon = 1/4$ would already improve the current best bound $O(p^{1/4} \log p)$ for the least quadratic nonresidue α_p, due to Burgess (1963), and the truth of Conjecture 1.3 would imply Linnik's conjecture that $\alpha_p << p^{\varepsilon}$ for all $\varepsilon > 0$. However it seems likely that improvements of Theorem 1.2 in the direction of (1.14) may be possible for small r, cf. the discussion at the end of §2.

Questions concerning the distribution of multiplicative dilations $a\mathsf{R} \pmod 1$ also arise in studying asymptotic denseness of sets on the torus R/Z, see Berend and Peres (1991). In particular Alon and Peres (1991) consider a closely related problem, concerning the size of the maximal gap in sets $a\mathsf{R} + b$ as a and b vary. They show that for every set $\mathsf{R} \pmod p$ there exists some $a \pmod p$ such that the set $a\mathsf{R}$ viewed on the circle $\mathsf{R}/p\mathsf{Z}$ has small discrepancy, i.e. all the intervals into which it cuts $\mathsf{R}/p\mathsf{Z}$ are of roughly the same length.

2. General Upper Bound

We use a second-moment method to establish the following bound.

Theorem 2.1. *Suppose that p is a prime and $\mathsf{R} = \{x_1, \dots, x_r\}$ is a set of integers with $1 \le x_1 < x_2 < \cdots < x_r \le p-1$. If $a \in \mathsf{Z}/p\mathsf{Z}$ is drawn with the uniform distribution, then*

$$\mathrm{Prob}\{\min[a\mathsf{R}] \ge \Delta\} \le \frac{1600p^2}{r\Delta^2} \tag{2.1}$$

holds for any positive Δ.

Proof. We use Fourier analysis on $\mathsf{Z}/p\mathsf{Z}$. Let $\chi_t(y)$ denote the characteristic function of $\{0, 1, \dots, t-1\}$, i.e.

$$\chi_t(y) = \begin{cases} 1 & 0 \le y \le t-1 , \\ 0 & t \le y \le p-1 . \end{cases}$$

Set $e(y) := \exp(\frac{2\pi iy}{p})$. Then χ_t has the Fourier series

$$\chi_t(y) = \sum_{k=0}^{p-1} a_k e(ky)$$

with coefficients

$$a_k = \frac{2}{p} \sum_{y=0}^{p-1} \chi_t(y) e(-ky) ,$$

and a simple calculation gives

$$a_k = \begin{cases} \frac{t}{p} & k = 0 , \\ \frac{\sin(\frac{\pi tk}{p})}{p \sin(\frac{\pi k}{p})} \, e(-\frac{(t-1)k}{2}) & 1 \le k \le p-1 . \end{cases}$$

We want a function whose Fourier coefficients drop off sufficiently rapidly in k and for this purpose use the convolution $f_t(y) = \chi_t * \chi_t(y)$ of $\chi_t(y)$ with itself. Recall that the convolution of two functions g and h is

$$g*h(y) = \frac{r}{p} \sum_{u=0}^{p-1} g(y-u) h(u)$$

and that Fourier coefficients of a convolution are the product of the Fourier coefficients of the factors. Hence

$$f_t(y) = \sum_{k=0}^{p-1} b_k e(ky)$$

has Fourier coefficients

$$b_k = \begin{cases} \dfrac{t^2}{p^2} & k = 0 , \\ \dfrac{\sin^2(\frac{\pi t k}{p})}{p^2 \sin^2(\frac{\pi k}{p})} e(-2(t-1)k) & 1 \le k \le p-1 . \end{cases} \tag{2.3}$$

The function f_t is nonnegative and is supported on the set $\{0, 1, 2, \ldots, 2t-2\}$. We will choose $t = \left\lceil \frac{\Delta}{2} \right\rceil$ which guarantees that f_t is supported on $\{0, 1, \ldots, \Delta\}$. Since only the case $\Delta \le p-1$ is of interest, we suppose that $t \le (p-1)/2$.

Now given the set R, we define the random variable

$$F_t(a) := \sum_{x \in R} f_t(ax) . \tag{2.4}$$

IF $F_t(a) \neq 0$ then $a\mathsf{R}$ must contain an element in the support of F_t, so that

$$\min[a\mathsf{R}] \le 2t-2 \le \Delta .$$

Hence

$$\text{Prob}\{\min[a\mathsf{R}] \ge \Delta\} \le \text{Prob}\{F_t(a) = 0\} . \tag{2.5}$$

It suffices to establish the upper bound (2.1) for $\text{Prob}\{F_t(a) = 0\}$ and for this we use Chebyshev's inequality, which asserts that any random variable F satisfies

$$\text{Prob}\{|F(a) - m| \ge \lambda\sigma\} \le \lambda^{-2} ,$$

where $m = E[F]$ and $\sigma^2 = E[F^2] - E[F]^2$ are its mean and variance, respectively. Applying this with $F = F_t$ and $\lambda = m/\sigma$ we obtain

$$\begin{aligned} \text{Prob}\{F_t(a) = 0\} &\le \text{Prob}\{|F_t(a) - E[F_t]| \le E[F_t]\} \\ &\le \frac{E[F_t^2] - E[F_t]^2}{E[F_t]^2} \ . \end{aligned} \tag{2.6}$$

To use this bound we calculate the first two moments of F_t.

The first moment $E[F_t]$ is easy to calculate. It is

$$\begin{aligned} E[F_t] &= \frac{1}{p} \sum_{a=0}^{p-1} F_t(a) \\ &= \frac{1}{p} \sum_{k=0}^{p-1} b_k \sum_{x \in \mathsf{R}} \left[\sum_{a=0}^{p-1} e(kax) \right] \\ &= rb_0 = r\frac{t^2}{p^2} \ , \end{aligned} \tag{2.7}$$

using (2.3).

It remains to obtain an upper bound for $E[F_t^2]$. To estimate it, we define

$$w(h,k) = |\{(x_1, x_2) : x_1, x_2 \in \mathsf{R} \text{ and } hx_1 \equiv kx_2 \pmod p\}| \ . \tag{2.8}$$

Then, since $F_t(x)$ is real,

$$\begin{aligned} E[F_t^2] &= \frac{1}{p} \sum_{a=0}^{p-1} F_t(a)^2 = \frac{1}{p} \sum_{a=0}^{p-1} F_t(a)\overline{F_t(a)} \\ &= \frac{1}{p} \sum_{a=0}^{p-1} \left[\sum_{x_1, x_2 \in \mathsf{R}} f_t(ax_1)\overline{f_t(ax_2)} \right] \\ &= \frac{1}{p} \sum_{a=0}^{p-1} \sum_{x_1, x_2 \in \mathsf{R}} \sum_{h,k=0}^{p-1} b_h \bar{b}_k e(a(hx_1 - kx_2)) \\ &= \frac{1}{p} \sum_{h,k=0}^{p-1} b_h \bar{b}_k \left\{ \sum_{x_1, x_2 \in \mathsf{R}} \sum_{a=0}^{p-1} e(a(bx_1 - kx_2)) \right\} \\ &= \sum_{h,k=0}^{p-1} b_h \bar{b}_k w(h,k) \ . \end{aligned}$$

We simplify this formula by observing that $w(0, 0) = r^2$, and $w(h,k) = 0$ if either $h = 0$ or $k = 0$ but not both. Thus we obtain

$$E[F_t^2] \;=\; r^2 b_0^2 \;+\; \sum_{h,k=1}^{p-1} b_h \overline{b}_k w(h,k) \;. \tag{2.9}$$

To bound this expression, we observe that

$$0 \le w(h,k) \le r$$

for all $(h,k) \ne (0,0)$, which gives

$$\Big| \sum_{h,k=1}^{p-1} b_h \overline{b}_k w(h,k) \Big| \le r \Big(\sum_{k=1}^{p-1} |b_k| \Big)^2 \;. \tag{2.10}$$

For the Fourier coefficients of F_t we have the trivial bound

$$|b_k| \le b_0 \;=\; \frac{t^2}{p^2}$$

and also the bounds

$$|b_k| \le \frac{4}{k^2} \qquad \text{if } 1 \le k < p/2 \;,$$
$$|b_k| \le \frac{4}{(p-k)^2} \qquad \text{if } p/2 < k \le p-1 \;,$$

which follow from (2.3) since $|\sin(x)| \ge \frac{2|x|}{\pi}$ for $|x| \le \pi/2$. These bounds imply that

$$\sum_{k=1}^{p-1} |b_k| \le 20 \frac{t}{p} \;, \tag{2.11}$$

on using the first bound above for the range $0 \le k \le \frac{p}{t}$ and the last two for the remainder. Combining (2.9)–(2.11) yields the second moment bound

$$E[F_t^2] \le \frac{r^2 t^4}{p^4} \;+\; 400 \frac{r t^2}{p^2} \;. \tag{2.12}$$

Substituting these first and second moment bounds into (2.6) yields

$$\text{Prob}\{F_t(a) = 0\} \le \frac{400p^2}{rt^2}$$
$$\le 1600\frac{p^2}{r\Delta^2},$$

since $t \ge \frac{\Delta}{2}$. ■

Theorem 1.2 is an immediate consequence of this result.

Proof of Theorem 1.2. Theorem 2.1 yields

$$E[\min[a\mathsf{R}]] \le \sum_{\Delta=0}^{p-1} \text{Prob}\{\min[a\mathsf{R}] \ge \Delta\}$$
$$\le \sum_{\Delta=1}^{p-1} \min\left[1, \frac{1600p^2}{r\Delta^2}\right] \le 100pr^{-1/2},$$

the desired bound. ■

To get stronger results in the direction of Conjecture 1.3 we must strengthen the bound for $\text{Prob}\{\min[a\mathsf{R}] \ge \Delta\}$. Examination of the proof of Theorem 2.1 suggests that better bounds than (2.1) are likely to hold, at least for certain ranges of r and Δ. Improvements might come by showing that the quantities $w(h,k)$ cannot be too large too often for small values of h and k, which are those smaller than $\frac{p}{\Delta^{1-\varepsilon}}$, where the products $|b_h b_k|$ are large. It is easy to see that

$$\sum_{h,k=1}^{p-1} w(h,k) = (p-1)r^2, \tag{2.13}$$

so that the quantities $w(h,k)$ are on average of size $\frac{r^2}{p-1}$. Some gain may be possible for r not too large, perhaps up to $r \le p^\beta$ for some $\beta < 1$. However for the quadratic nonresidue example N_p one has $r = \frac{p-1}{2}$ and improvements in bounding (2.9) appear to require cancellation involving the complex arguments of the Fourier coefficients b_k.

In addition, the use of the second moment method itself presumably gives something away,

because Chebyshev's inequality is not tight for distributions having smooth tails. Estimates for higher moments might conceivably yield improved bounds for Prob$\{\min[aR] \geq \Delta\}$. Such moment estimates involve various interesting problems concerning the distribution of solutions to linear Diophantine equations (mod p) with bounds on the variables. In Theorem 2.1 they concern the ensemble of quantities $w(h,k)$, and other examples of such problems appear in Alon and Peres (1991), Lemma 5.1, and in Lagarias and Håstad (1986).

3. Constant $|R|$ Case: Two Examples

Now we consider $M^*(R)$ for $|R| = r$ of a fixed size as $p \to \infty$. We estimate $M^*(R)$ for the sets $I_r = [1, 2, \ldots, r]$ and $J_{2r} = [\pm 1, \pm 2, \ldots, \pm r]$, which give relatively large and relatively small values of $M^*(R)$, respectively.

Theorem 3.1. *One has*

$$M^*(I_r) = \frac{1}{4} \sum_{\frac{h}{k}, \frac{h'}{k'} \text{ adj in } \mathcal{F}_r} \frac{1}{kk'}\left(\frac{1}{k} + \frac{1}{k'}\right) + O\left(\frac{r}{p}\right), \tag{3.1}$$

as $p \to \infty$, where the sum runs over all intervals $(\frac{h}{k}, \frac{h'}{k'})$ in the Farey series $\mathcal{F}_r$ of order r.

Proof. Divide the real interval $[0, p]$ into segments $[\frac{h}{k}p, \frac{h'}{k'}p]$ using the dissection $p\mathcal{F}_r$.

Claim. *For those $a \in (\mathbb{Z}/p\mathbb{Z})^*$ for which $\frac{h}{k}p \leq a < \frac{h'}{k'}p$, the minimal element in aI_r (mod p) is ak (mod p).*

Proof of claim. Consider the piecewise linear function $f(x) = kx - p[\frac{kx}{p}]$, written $f(x) = kx \pmod p$, on the interval $[\frac{h}{k}p, \frac{h'}{k'}p]$. It is 0 at the left endpoint $\frac{h}{k}p$, and it is $\frac{1}{k'}p$ at the right endpoint $\frac{h'}{k'}p$ since the interval has length $\frac{p}{kk'}$. If some $f(x) = hx \pmod p$ with $0 < h < r$, $h \neq k$ were smaller than it anywhere inside the interval, then it must have an intersection point inside the interval. Any such intersection point satisfies $kx = hx + ap$ for some $|a| < r$, hence

$x = |\frac{a}{k-<}|p$ and $|\frac{a}{k-<}| \in \wedge_r$ with $\frac{<}{k} < |\frac{a}{k-l}| < \frac{<'}{k'}$, a contradiction proving the claim. ■

Now define

$$s(\frac{<}{k}, \frac{<'}{k'}) := \sum_{\frac{<}{k}p \le a \le \frac{<'}{k'}p} \min(a|_r) .$$

By the claim,

$$s(\frac{<}{k}, \frac{<'}{k'}) = \sum_{\frac{<}{k}p \le a \le \frac{<'}{k'}p} ak(\mathrm{mod}\ p) .$$

This gives

$$s(\frac{<}{k}, \frac{<'}{k'}) = \frac{1}{2}\frac{p^2 k}{(kk')^2} + O(\frac{p}{k'}) . \tag{3.2}$$

The main term in this expression is the area under the line kx (mod p) in the interval, see Figure 3.1.

Insert Figure 3.1 about here.

Using this estimate

$$\begin{aligned}(p(p-1)M^*(|_r) &= \sum_{\wedge_r} s(\frac{<}{k}, \frac{<'}{k'}) \\ &= \frac{p^2}{2}\sum_{\wedge_r}\frac{1}{kk'}(\frac{1}{k'}) + O(p\sum_{\wedge_r}\frac{1}{k'}) .\end{aligned} \tag{3.3}$$

Now use the fact that the Farey series is symmetric about 1/2, hence if $(\frac{<}{k}, \frac{<'}{k'}) \in \wedge_r$ then $\left[\frac{k'-l'}{k'}, \frac{k-l}{k}\right] \in \wedge_r$. Pairing these terms gives

$$\sum_{\wedge_r} \frac{1}{kk'}\left(\frac{1}{k'}\right) = \frac{1}{2}\sum_{\wedge_r} \frac{1}{kk'}\left(\frac{1}{k} + \frac{1}{k'}\right) .$$

Now (3.1) follows by dividing (3.3) by $p(p-1)$. There is a remainder arising from both terms on the right side of (3.3), and it is bounded using

$$\sum_{\wedge_r} \frac{1}{k'} \le \sum_{k'=2}^{r} \frac{1}{k'}\left(\sum_{j=1}^{k'} 1\right) + 2 = r + 1 , \tag{3.4}$$

which gives the result. ■

Now define

$$D_r := \sum_{\wedge_r} \frac{1}{kk'}\left(\frac{1}{k} + \frac{1}{k'}\right) .$$

The sums D_r were studied by Hans and Chander (1964) (see also Robertson (1968)), who showed that

$$D_r = \left(\frac{\pi^2}{6} + o(1)\right) \frac{r}{\log r} \tag{3.5}$$

as $r \to \infty$. In particular $D_r \ge c_0 \frac{r}{\log r}$ for some absolute constant $c_0 > 0$ for all r. This yields:

Corollary 3.1a. *One has*

$$M^*(\mathrm{I}_r) = \frac{1}{4} D_r + O\left(\frac{r}{p}\right) \tag{3.6}$$

as $p \to \infty$, *where* $D_r = \frac{\pi^2}{6}(1 + o(1))$ *as* $r \to \infty$.

Next we treat $\mathrm{J}_{2r} = [\pm 1, \pm 2, \ldots, \pm 2r]$.

Theorem 3.2. *For fixed r and* $p \to \infty$,

$$M(\mathrm{J}_{2r}^*) = \frac{1}{2} \sum_{\wedge_r} \frac{1}{kk'}\left(\frac{1}{k + k'}\right) + O\left(\frac{r}{p}\right) , \tag{3.7}$$

where $\left(\frac{<}{k}, \frac{<'}{k'}\right)$ *runs over all intervals of the Farey series* $\wedge_r$ *of order r.*

Proof. We proceed similarly to Theorem 3.1.

Claim. *For those* $a \in \mathbb{Z}/p\mathbb{Z}$ *with* $\frac{<}{k}p \le a \le \frac{<'}{k'}p$ *the minimal element in* $a\mathcal{J}_{2r}^{*}$ *(mod p) is* ak *(mod p) for* $\frac{<}{k}p < a \le (\frac{<}{k} + \frac{1}{k(k+k')})p$ *and* $-ak'$ *(mod p) for* $(\frac{<}{k} + \frac{1}{k(k+k')})p \le a \le \frac{<'}{k'}p$.

Proof of claim. The proof of Theorem 3.1 showed that in the interval $(\frac{<}{k}p, \frac{<'}{k'}p)$ the function ak (mod p) lies below all $a_<$ (mod p) with $< \ne k$ for $1 \le < \le r$. A similar argument shows that the function $-ak'$ (mod p) lies below all $-a_<$ (mod p) with $< \ne k'$ for $1 \le < \le r$ on the interval. Hence the minimal value for $a\mathcal{J}_{2r}^{*}$ is $\min(ak, -ak')$ on the interval, and determining which is smaller gives the stated result. ■

Now set

$$s^{*}(\frac{<}{k}, \frac{<'}{k'}) = \sum_{\frac{<}{k}p \le a \le \frac{<'}{k'}p} \min(a\mathcal{J}_{2r}) \ .$$

Using the claim, we have

$$\begin{aligned} s^{*}(\frac{<}{k}, \frac{<'}{k'}) &= \sum_{\frac{<}{k}p < a \le \frac{<'}{k'}p} \min(ak, -ak') \pmod{p} \\ &= \frac{p^2}{2}\ \frac{1}{kk'}(\frac{1}{k+k'}) + O(\frac{p}{k+k'}) \ , \end{aligned} \tag{3.8}$$

where the main term in this expression is the area of the triangle pictured in Figure 3.2.

Insert Figure 3.2 about here.

Proceeding as in Theorem 3.1, we obtain

$$p(p-1)M^*(\mathrm{J}_{2r}) = \frac{p^2}{2} \sum_{\wedge_r} \frac{1}{kk'}\left(\frac{1}{k} + \frac{1}{k'}\right) + O\left(p \sum_{\wedge_r} \frac{1}{k+k'}\right),$$

and dividing by $p(p-1)$ yields (3.7). ■

Now set

$$E_r := \sum_{\wedge_r} \frac{1}{kk'}\left(\frac{1}{k+k'}\right).$$

These sums were estimated asymptotically by Lehner and Newman (1969), Theorem 2, who showed that

$$E_r = \frac{12 \log 2}{\pi^2} \frac{1}{r} + O\left(\frac{\log r}{r^2}\right). \tag{3.9}$$

This yields:

Corollary 3.2a. *One has*

$$M^*(J_{2r}) = \frac{1}{2}E_r + O\left(\frac{r}{p}\right) \quad \text{as } p \to \infty,$$

where $E_r = \frac{12 \log 2}{\pi^2} \frac{1}{r} + O\left(\frac{\log r}{r^2}\right)$ *as* $r \to \infty$.

Acknowledgements. We are indebted to Gary Miller and Sandeep Sen for bringing problems of this kind to our attention and to Yuval Peres for pointing out the quadratic residue example in [4].

References

[1] M. Ajtai, J. Komlos, E. Szemeredi (1990), Generating expanders from two permutations; in: *A Tribute to Paul Erdös* (A. Baker, B. Bollobas, A. Hajnal, Eds.), Cambridge U. Press, 1-12.

[2] N. Alon, L. Babai and A. Itai (1986), A fast and simple randomized parallel algorithm for the maximal independent set problem, J. Algorithms 7, 567-583.

[3] N. Alon and Y. Peres (1991), Uniform dilations, preprint.

[4] D. Berend and Y. Peres (1991), Asymptotically dense dilations of sets on the circle, J. London Math. Soc., to appear.

[5] D. A. Burgess (1963), A note on the distribution of residues and nonresidues, J. London Math. Soc. **38,** 253-256.

[6] S. W. Graham and C. J. Ringrose (1990), Lower bounds for least quadratic nonresidues, in: *Analytic Number Theory: Proceedings of a Conference in Honor of P. Bateman* (B. Berndt et al., eds.), Birkhäuser, Boston.

[7] R. J. Hans and V. Chander (1964), An interesting identity, Res. Bull. Panjab Univ. **15,** 353-356.

[8] G. H. Hardy and E. M. Wright (1960), *Introduction to the Theory of Numbers* (4^{th} Edition), Oxford U. Press.

[9] A. Joffe (1974), On a set of almost deterministic *k*-independent random variables, Ann. Prob. **2,** 161-162.

[10] J. C. Lagarias (1990), Pseudorandom number generators in cryptography and number theory, in: *Cryptology and Computational Number Theory*, C. Pomerance, Ed., Proc.

Symp. Applied Math. 42, American Math. Society, 115-143.

[11] J. C. Lagarias and J. Håstad (1986), Simultaneous diophantine approximation of rationals by rationals, J. Number Theory **24,** 200-228.

[12] J. Lehner and M. Newman (1969), Sums involving Farey fractions, Acta Arithmetica **15,** 182-187.

[13] M. Luby (1985), A simple parallel algorithm for the maximal independent set problem, Proc. 11th ACM Symp. on Theory of Computing, ACM Press, 1-10.

[14] M. M. Robertson (1968), Sums associated with Farey series, Prob. Camb. Phil. Soc. **64,** 393-398.

[15] D. Zuckerman (1990), General weak random sources, Proc. 21st IEEE Conf. on Foundations of Computer Science, Vol. II, 534-543.

On the Distribution of Multiplicative Translates of Sets of Residues (mod p)

J. Håstad

Royal Institute of Technology
Stockholm, Sweden

J. C. Lagarias

A. M. Odlyzko

AT&T Bell Laboratories
Murray Hill, NJ 07974

(July 29, 1992)

ABSTRACT

Let R be a set of r distinct nonzero residues modulo a prime p, and suppose that the random variable a is drawn with the uniform distribution from $\{1, 2, \ldots, p-1\}$. We show for all sets R that $\frac{p-2}{2r} \le E[\min[a\mathsf{R}]] \le 100\frac{p}{r^{1/2}}$, where in the set $a\mathsf{R}$ each integer is identified with its least positive residue modulo p. We give examples where $E[\min[a\mathsf{R}]] \le \frac{0.8p}{r}$ and $E[\min[a\mathsf{R}]] \ge 0.4\ \frac{p \log r}{r}$. We conjecture that $E[\min[a\mathsf{R}]] << \frac{p}{r^{1-\varepsilon}}$ holds for a wide range of r. These results are applicable to the analysis of certain randomization procedures.