On the Stochasticity Parameter of Quadratic Residues

M. R. Gabdullin

Presented by Academician of the RAS S.V. Konyagin November 12, 2019

Received November 14, 2019; revised November 20, 2019; accepted December 12, 2019

Abstract—Following V.I. Arnold, we define the stochasticity parameter \( S(U) \) of a set \( U \subseteq \mathbb{Z}_M \) to be the sum of squares of consecutive distances between the elements of \( U \). The stochasticity parameter of the set \( R_M \) of quadratic residues modulo \( M \) is studied. We compare \( S(R_M) \) with the average value \( s(k) = s(k, M) \) of \( S(U) \) over all subsets of \( U \) of size \( k \). It is proved that (a) for a set of moduli of positive lower density, we have \( S(R_M) < s(|R_M|) \); and (b) for infinitely many moduli, \( S(R_M) > s(|R_M|) \).

Keywords: quadratic residues, stochasticity parameter

DOI: 10.1134/S106456242002012X

Consider an arbitrary subset \( U \) of a residue ring \( \mathbb{Z}_n \). Suppose that
\[ U = \{0 \leq u_1 < u_2 < \ldots < u_k < n\} \]
and \( u_{k+1} = n + u_l \) (i.e., \( \{0, n\} \) is “glued” in a circle). To examine the randomness of the point distribution of \( U \) over all residues, Arnold introduced the stochasticity parameter
\[ S(U) = \sum_{i=1}^{k} (u_{i+1} - u_i)^2 \]
(see [1, Section 9]). It is easy to see that \( S(U) \) is minimal when the points are at identical intervals and is maximal when all of them are in one place. Therefore, extremely small or extremely large values of \( S(U) \) suggest that the behavior of \( U \) is nonrandom. Keeping in mind these two extreme cases, we say that the points of \( U \) exhibit repulsion (or attraction) if \( S(U) \) is less than (greater than, respectively) the average value of the stochasticity parameter over all subsets of \( \mathbb{Z}_n \) of size \( k \).

Garaev, Konyagin, and Malykhin [2] generalized the stochasticity parameter. Given \( q > 0 \), define
\[ S_q(U) = \sum_{i=1}^{k} (u_{i+1} - u_i)^q. \]
Let \( N_l(U) \), \( l \geq 1 \), denote the number of gaps of length \( l \) in the set \( U \), i.e., the number of elements \( x \in \mathbb{Z}_n \) such that \( x \in U, x + 1, \ldots, x + l - 1 \notin U \), and \( x + l \in U \). Then
\[ S_q(U) = \sum_{l=1}^{\infty} N_l(U) l^q \]
(1)
in particular, \( S(U) = \sum_{l=1}^{\infty} N_l(U) l^2 \). The authors prove that, if \( |U| = k < n \) and \( \frac{k}{\sqrt{n}} \to \infty \), then the average value \( s_q(k, n) \) of \( S_q(U) \) for a set \( U \) of size \( k \) can be represented in the form
\[ s_q(k, n) = \frac{k^2}{n} \sum_{l=1}^{\infty} \left(1 - \frac{k}{n}\right)^l l^q (1 + o(1)). \]
Note that this asymptotic formula agrees with the following heuristic considerations: the probability of a point of a random set of size \( k \) being the beginning of a gap of length \( l \) is equal to \( \frac{k^2}{n} \left(1 - \frac{k}{n}\right)^l \), so, on average, \( N_l(U) \) must be close to \( \frac{k^2}{n} \left(1 - \frac{k}{n}\right)^l \).

Additionally, the authors noted that, in the case \( q = 2 \), the quantity \( s(k) := S_2(k, M) \) can be written in explicit form.

Proposition 1. It is true that
\[ s(k) = \frac{M(2M - k + 1)}{k + 1}. \]
In [2] \( S_q(G) \) was estimated in the case where \( n = p \) is a prime number and the set \( G \) is a subgroup \( G_r \subset \mathbb{Z}_p^* \) of order \( t \), i.e., consists of the \( d \)th powers of the num-
bers from 1 to \( p - 1 \), where \( d = \frac{p-1}{t} \). Below is the main result of [2].

**Theorem A.** There exists a constant \( c > 0 \) such that, for \( q \in (0, 4) \) and \( d \leq \exp(c\sqrt{\log p}) \), the following asymptotic formula is valid:

\[
S_q(G_t) = S_q\left(\frac{p-1}{d}\right)(1 + o(1)), \quad p \to \infty.
\]

Thus, in terms of the behavior of \( S_q(G_t) \), large multiplicative subgroups of \( \mathbb{Z}_p^* \) are close to random sets of the corresponding size.

In this work, we study the stochasticity parameter for the set \( R_M \) of quadratic residues modulo \( M \). We prove the repulsion of quadratic residues for a set of moduli of positive lower density (here and in what follows, we mean the lower asymptotic density, i.e., \( d(A) = \lim \inf_{N \to \infty} \frac{|A \cap [1, \ldots, N]|}{N}, A \subseteq \mathbb{N} \)) and prove their attraction for infinitely many moduli. Now we give rigorous formulations. Let \( c_0 \) and \( C_0 \) be absolute positive constants, where \( c_0 \) is sufficiently small and \( C_0 \) is sufficiently large. Denote by \( \Omega \) the set of numbers \( M \) such that \( M = Am \), where \( m = p_1 \ldots p_t \geq 0.4 \log \log M, \quad p_1 < p_2 < \ldots < p_t \) are distinct primes greater than \( 2^{c_0} \), and \( A \) is a square-free number such that \( (A, m) = 1 \) and \( A \leq 2^{c_0} \).

Below is the main result of this paper.

**Theorem 1.** There exists an absolute constant \( c > 0 \) such that, for \( M \in \Omega \), we have the asymptotic formula

\[
S(R_M) = m2^c + A^2 |R_A|^{-1} - A^2 |R_A|^{-1} m + E,
\]

where

\[
E \ll m2^c A^2 p_1^{-1} + m A^2 |R_A|^{-1} 2^c = o(m), \quad M \to \infty, \quad M \in \Omega.
\]

Moreover, the set \( \Omega \) has a lower positive density.

On the other hand, for moduli \( M \in \Omega \), Proposition 1 yields

\[
s(|R_M|) = m2^c A^2 |R_A|^{-1} - M + O(A^2 |R_A|^{-1} m2^c p_1^{-1}).
\]

Thus, the leading terms in the asymptotic formulas for \( S(R_M) \) and \( s(|R_M|) \) coincide. From this, we obtain an analogue of Theorem A.

**Corollary 1.** It is true that

\[
S(R_M) = s(|R_M|)(1 + o(1)), \quad M \to \infty, \quad M \in \Omega.
\]

Comparing the second terms in (2) and (3), we observe the repulsion of quadratic residues.

**Corollary 2.** For all sufficiently large \( M \in \Omega \) with \( A \geq 3 \),

\[
S(R_M) < s(|R_M|).
\]

Additionally, for moduli of the form \( M = Ap \), where \( p \) is a prime and \( (A, p) = 1 \), we can derive an asymptotic formula with a smaller remainder, which makes it possible to establish the attraction of quadratic residues for infinitely many moduli.

**Theorem 2.** Let \( M = Ap \) and \( (A, p) = 1 \). Then

\[
S(R_M) = 2f_A(0.5)p + O(A^2 p^{-c}),
\]

where \( f_A \) is a rational function determined by the number \( A \).

The coefficients of \( f_A \) can be written in explicit form. Let \( s_1, \ldots, s_{|R_A|} \) be the consecutive distances between the quadratic residues modulo \( A \) indexed by the residues modulo \( |R_A| \). Then

\[
f_A(y) = \frac{f(y)}{Q(y)},
\]

where \( Q(y) = Q_A(y) = 1 + y + \ldots + y^{|R_A|-1} \) and \( f(y) = f_A(y) = \sum_{k=0}^{|R_A|} \beta_k y^k \) is a reciprocal polynomial with coefficients \( \beta_i = \beta_{|R_A|} = \sum_{j=0}^{|R_A|} s_j^{2i} \) and \( \beta_k = 2\sum_{j=0}^{|R_A|} s_j s_{i+k} \) for \( 0 < k < |R_A| \) (we assume that the indices \( i \) of the numbers \( s_j \) are calculated modulo \( |R_A| \)). Note that the behavior of \( f_A \) near the point \( y = 1 \) is a determining factor for the attraction or repulsion of quadratic residues.

For moduli of the same form, Proposition 1 (or formula (3) with \( t = 1 \)) yields

\[
s(|R_M|) = \left(\frac{4A^2}{|R_A|} - A\right)p + O(A^2 |R_A|^{-1}).
\]

Numerical check shows that

\[
2f_A(0.5) < \frac{4A^2}{|R_A|} - A
\]

for all \( 3 \leq A \leq 100 \) with \( A \neq 89 \) and

\[
2f_A(0.5) > \frac{4A^2}{|R_A|} - A
\]

for \( A = 89 \). Thus, Theorem 2 implies the following result.

**Corollary 3.** It is true that

\[
\lim_{M \to \infty} \frac{S(R_M)}{s(|R_M|)} < 1 < \lim_{M \to \infty} \frac{S(R_M)}{s(|R_M|)}.
\]

It follows from the Cauchy–Schwarz inequality that

\[
\lim_{M \to \infty} \frac{S(R_M)}{s(|R_M|)} \geq 0.5.
\]

At the same time, we do not know whether the upper limit \( \lim_{M \to \infty} \frac{S(R_M)}{s(|R_M|)} \) is finite. Note that Aryan [3] proved that, for square-free \( M \),

DOKLADY MATHEMATICS  Vol. 101  No. 2  2020
Let us discuss the basic ideas underlying the proof of Theorem 1 (the proof of Theorem 2 is similar, but technically somewhat easier). Fixing \( \alpha \in \left(0, \frac{1}{10}\right) \) and a sufficiently small number \( \varepsilon_0 > 0 \), we define \( d_1 := \frac{\alpha A}{2|R_A|} \log p_1 \) and \( d_2 := A^\frac{\varepsilon_0}{|R_A|} \). The quantity \( S(R_M) \) is written as \( \sum_{j=2} N_j(R_M)^2 \) (see (1)) and this sum is split into three parts. Next, we find an asymptotic formula for the contribution made by small gaps of length \( l \leq d \) in terms of the function \( f_A \) determined by the number \( A \) and show that the contributions made by medium gaps (of length \( l \in (d_1, d_2) \)) and large gaps (of length \( l > d_2 \)) are asymptotically small. More precisely, we prove that

\[
S(R_M) = m \cdot 2^{\ell} f_A(y_\ell) + O(m \cdot 2^{\ell} A^4 p_1^{-c_1}),
\]

where \( c_1 > 0 \) is an absolute constant and \( y_\ell = 1 - 2^{-\ell} \). Finally, we show that \( f_A(l) = 2A^2|R_M|^{-1}, f_A'(l) = A^2|R_A|^{-1}, \) and \( f_A''(y) \ll A^2|R_A|^{-1} \) for \( y \in (y_\ell, 1) \), assuming that \( |R_A| \ll 2^\ell \). Then the first assertion of Theorem 1 is derived by expanding \( f_A \) in a Taylor series around the point \( y = 1 \). Finally, the fact that \( \Omega \) has a positive lower density is proved by applying combinatorial arguments and sieve methods.

Note that an asymptotic formula for the contribution made by small gaps can be derived using the asymptotic distribution of gaps between quadratic residues found by Kurlberg and Rudnick [4] for square-free moduli and by Kurlberg [5] for arbitrary moduli. Let \( 0 = u_0 < u_1 < \ldots < u_{R_0} < M \) be the set of quadratic residues modulo \( M \). The following result was obtained in [5].

**Theorem B.** For \( t \geq 0, \)

\[
|R_M|^{-1} \sum_{j \in \left\{1, \ldots, |R_M|\right\}} u_j - u_{j-1} > tM|R_M|^{-1} \leq e^{-t} + o(1)
\]

as \( \alpha(M) \rightarrow \infty \); moreover, for any fixed \( t_0 > 0 \), this bound is uniform for \( t \in [0, t_0] \).

Nevertheless, the term \( o(1) \) on the right-hand side prevent us from finding the second term in the asymptotics for \( S(R_M) \), so we rely on other considerations (and consider moduli of a special form).

The proof of Theorem 1 also makes it possible to obtain a lower bound (rather weak) on the density of \( \Omega \) (writing it in explicit form or optimizing was beyond the scope of this work).

Apparently, the method described fails to capture attraction for a large number of moduli. Nevertheless, it seems plausible that the attraction of quadratic residues holds for a set of moduli of positive lower density.

**Finding**

This work was supported by the Russian Science Foundation, project no. 19-11-00001.

**References**

1. V. I. Arnold, *Euler Groups and Arithmetics of Geometric Progressions* (MTsNMO, Moscow, 2003) [in Russian].
2. M. Z. Garaev, S. V. Konyagin, and Yu. V. Malykhin, Proc. Steklov Inst. Math. 276, 77–89 (2012).
3. F. Aryan, Int. Math. Res. Not. IMRN 23, 12405–12431 (2015).
4. P. Kurlberg and Z. Rudnick, Duke Math. J. 100, 211–242 (1999).
5. P. Kurlberg, Isr. J. Math. A 120, 205–224 (2000).

Translated by I. Ruzanova