On the measure of equidistribution of point sets.

By ALFRÉD RÉNYI in Budapest.

Introduction.

Throughout the paper we are concerned with measurable point sets E lying in the interval $(0, 1)$. The measure of E shall be denoted by $|E|$ and the characteristic function of E by $F(x)$. We define $F(x)$ outside the interval $(0, 1)$ so as to be periodic with period 1. We denote by E_t (for any real t) the set which has the characteristic function $F(x+t)$. If we imagine the interval $(0, 1)$ wound on a circle of circumference unity, we may say that E_t is obtained by rotating the set E by the angle $-t$. Let $G(t)$ denote the measure of the set of points of the interval $(0,1)$ which are common to E and E_t. We have evidently

$$G(t)=\int_0^1 F(x)\,F(x+t)\,dx. \tag{1}$$

$G(t)$ is a non-negative function, periodic with period 1. We have, in view of the periodicity of $F(x)$,

$$G(t)=\int_0^1 F\left(x-\frac{t}{2}\right)F\left(x+\frac{t}{2}\right)dx, \tag{2}$$

thus $G(t)$ is an even function. Further we have

$$|G(t+h)-G(t)|\leqq\int_0^1 |F(x+h)-F(x)|\,dx. \tag{3}$$

Now, it is well known[1]) that the integral on the right side tends to 0 with h, thus $G(t)$ is continuous. As we have $G(0)=|E|$, it follows from the continuity of $G(t)$ that if $|E|>0$, there exists a constant $c>0$, for which $G(t)>0$ for $0\leqq t<c$. This is equivalent to a theorem of

[1]) Cf. for ex. A. ZYGMUND, *Trigonometrical series* (Warszawa, 1935), p. 17.

H. STEINHAUS[2]), who stated it in the form, that the set of the mutual distances of the points of a set of positive measure contains a whole interval $(0, c)$. In view of this interpretation, we shall call $G(t)$ the *distance function* of the set E.

Now let us denote the minimal value of the continuous function $G(t)$ by $m(E)$. As $G(t)$ is non-negative, further as we have

$$\int_0^1 G(t)\,dt = \int_0^1\int_0^1 F(x)\,F(x+t)\,dx\,dt = |E|^2, \tag{4}$$

it follows

$$0 \leqq m(E) \leqq |E|^2. \tag{5}$$

It is easy to see that $m(E) = |E|^2$ if and only if $|E| = 0$ or $|E| = 1$. Thus if we put

$$\mu(E) = \frac{m(E)}{|E|^2}, \tag{6}$$

we have $0 \leqq \mu(E) < 1$ for $0 < |E| < 1$. In what follows $\mu(E)$ shall be called the *measure of equidistribution* of the set E. Of course the notion of equidistribution, implied by this definition, is different from (but as we shall see is closely connected with) the usual definition for sequences, introduced by H. WEYL[3]). The difference is made clear by remarking that we are concerned not with the equidistribution of the *points* of E but with the equidistribution of the set of *distances* of pairs of points of E.

The purpose of the present paper is to prove that there exist sets having any prescribed positive measure, and as "highly equidistributed" as we please, i. e., having a measure of equidistribution arbitrarily near to **1**. This shall be proved in **§. 2** (Theorem 1). **§. 1** contains preliminary discussions of rather general character, concerning the FOURIER expansion of the distance function and some lemmas. The proof of Theorem 1 is based on a property of quadratic residues, discovered by LAGRANGE[4]). In **§. 3** the problem is generalized. We introduce the notion of the measure of k-fold equidistribution, and prove a theorem, analogous to, but somewhat weaker than Theorem 1 (Theorem 2), based

---

[2]) H. STEINHAUS, Sur les distances des points des ensembles de mesure positive, *Fundamenta Math.*, **1** (1920), pp. 93—104. Cf. also S. PICCARD, Sur les ensembles de distances des ensembles de points d'un espace euclidien, *Mémoires Université Neuchatel*, **13** (1939), pp. 212.

[3]) H. WEYL, Über die Gleichverteilung von Zahlen mod. Eins, *Math. Annalen*, **77** (1916), pp. 313—325.

[4]) P. BACHMANN, *Niedere Zahlentheorie*, Vol. II, (Leipzig, 1910), pp. 241—245.

on a generalization of a theorem of THUE[5]). In §. 4, we point out the connection with some problems of number theory, and prove a theorem concerning the sequences of integers, called difference bases, constructed by SINGER[6]) (Theorem 3).

§. 1. Fourier expansion of the distance function.

Let $F(x)$ denote the characteristic function of a measurable set E in the interval $(0,1)$. Let us consider the FOURIER expansion of $F(x)$:

$$F(x) \sim a_0 + 2\sum_{n=1}^{\infty} (a_n \cos 2\pi n x + b_n \sin 2\pi n x).$$

Lemma 1. *If $G(t)$ denotes the distance function of the set E as defined in the introduction, we have*

$$G(t) = a_0^2 + 2\sum_{n=1}^{\infty} (a_n^2 + b_n^2) \cos 2\pi n t.$$

The series on the right converges uniformly.

Evidently Lemma 1 follows from Parseval's theorem.

In what follows we shall consider some special sets consisting of a finite number of intervals of equal length. Let $b_0, b_1, \ldots, b_{N-1}$ denote a sequence of integers, which are all different modulo q. The set $E = E_q(b_0, b_1, \ldots, b_{N-1})$ shall be defined as the set consisting of the intervals $\left(\frac{b_j - \frac{1}{2}}{q}, \frac{b_j + \frac{1}{2}}{q}\right)$, $(j = 0, 1, \ldots, N-1)$. Evidently, the set E is not changed if one of the b_j is replaced by a number congruent to it modulo q, thus we may suppose $0 \leqq b_j < q$.

Lemma 2. *Let $G(t)$ denote the distance function of a set $E = E_q(b_0, b_1, \ldots, b_{N-1})$. Let us denote*

$$c_n = \sum_{j=0}^{N-1} \exp\left(\frac{2\pi i n b_j}{q}\right). \tag{7}$$

Then we have

$$G(t) = \frac{N^2}{q^2} + \frac{2}{q^2}\sum_{n=1}^{\infty} |c_n|^2 \left(\frac{\sin\frac{\pi n}{q}}{\frac{\pi n}{q}}\right)^2 \cos 2\pi n t. \tag{8}$$

---

[5]) A. SCHOLZ, *Einführung in die Zahlentheorie* (Sammlung Göschen, Bd. 1131, Berlin, 1939), p. 45.

[6]) I. SINGER, A theorem in finite projective geometry and some applications to number theory, *Transactions American Math. Society*, 43 (1938), pp. 377–385. Cf. also: T. VIJAYARAGHAVAN and S. CHOWLA, Short proof of theorems of Bose and Singer, *Proceedings National Academy Sciences India*, Section A, 15 (1945), p. 194.

Lemma 2 is verified easily by calculating explicitly the FOURIER coefficients of the characteristic function of the set E and applying Lemma 1.

Lemma 3. *Let us have* $0<h<1$. *We define*

$$R_h(x)=h^2+2h^2\sum_{n=1}^{\infty}\left(\frac{\sin n\pi h}{n\pi h}\right)^2\cos 2\pi n x. \tag{9}$$

Then we have

$$R_h(x)=h-|x| \quad \text{for} \quad |x|\leqq h,$$
$$R_h(x)=0 \quad \text{for} \quad |x|>h.$$

Lemma 3 is easily verified by calculating the FOURIER coefficients of $R_h(x)$. The function $R_h(x)$ may be called the "RIEMANN kernel". As a matter of fact, $f(x)$ denoting a function, L-integrable in (0, 1), the summation method of RIEMANN consists in forming the second generalized derivative of the function $\psi(x)$, obtained by integrating $f(x)$ twice, and it is easy to see that we have

$$\frac{\psi(x+2h)+\psi(x-2h)-2\psi(x)}{4h^2}=\frac{1}{h^2}\int_0^1 f(t)\,R_h(x-t)\,dt, \tag{10}$$

i. e. $R_h(x)$ is the kernel function of the RIEMANN summation[7]).

It can be seen from (7) that $c_n=c_m$ if $n\equiv m \bmod q$. Further, as $\sin\frac{n\pi}{q}=0$ for $n\equiv 0 \bmod q$, the values of c_n for $n\equiv 0 \bmod q$ figure in the expansion (8) only formally, and the FOURIER expansion of $G(t)$ is completely determined if we know the values of $c_1, c_2, \ldots, c_{q-1}$. Lemma 3 shows that $G(t)$ can easily be calculated if the values of $|c_n|^2$ $(n\not\equiv 0)$ are all equal. The same is true if they show only relatively small deviations from a common value. This is expressed by the following

Lemma 4. *Let* $E=E_q(b_0, b_1, \ldots, b_{N-1})$ *be defined as above. If the numbers* c_n *defined by* (7) *satisfy the relations*

$$\left||c_n|^2-Q\right|<\frac{\vartheta Q}{q-1} \quad \text{for} \quad n=1, 2, \ldots, q-1,$$

where $Q(1+\vartheta)<N^2$, *we have*

$$\mu(E)\geqq 1-\frac{Q(1+\vartheta)}{N^2}.$$

Proof. We have evidently from (8)

$$G(t)\geqq\frac{N^2}{q^2}+Q\left(R_{1/q}(t)-\frac{1}{q^2}\right)-\frac{\vartheta Q}{q-1}\left(R_{1/q}(0)-\frac{1}{q^2}\right) \tag{11}$$

---

[7]) This has been already remarked by M. SCHECHTER, Über die Summation divergenter Fourier-Reihen, *Monatshefte für Math. und Physik*, 25 (1911), pp. 224–234. It was Prof. L. FEJÉR who has kindly called my attention to this paper.

and thus

$$m(E) \geqq \frac{N^2 - Q(1+\vartheta)}{q^2} \tag{12}$$

and Lemma 4 follows easily.

Of course the situation is the simplest if, in Lemma 4, $\vartheta = 0$. Sequences of integers b_j for which this holds, are characterized by the following

Lemma 5. *If $b_0, b_1, b_2, \ldots, b_{N-1}$ denote a sequence of integers with the property that the differences $b_r - b_s$ $(r, s = 0, 1, \ldots, N-1; r \neq s)$ represent every class of residues modulo q (the class 0 of course excepted) exac ly k-times, we shall call the sequence b_j a difference basis of order k modulo q. The necessary and sufficient condition for the sequence b_j being a difference basis of order k modulo q, is that for any $n \not\equiv 0 \bmod q$*

$$\left| \sum_{r=0}^{N-1} \exp\left(2\pi i \frac{b_r n}{q}\right) \right|^2 = N - k \tag{13}$$

be valid.

It is clear that the condition (13) is necessary. Let us prove that it is also sufficient. Let A_l $(l = 1, 2, \ldots, q-1)$ denote the number of representations of l mod q in the form $b_r - b_s$. We have

$$N - k = \left| \sum_{r=0}^{N-1} \exp\left(2\pi i \frac{b_r n}{q}\right) \right|^2 = N + \sum_{l=1}^{q-1} A_l \exp\left(2\pi i \frac{ln}{q}\right).$$

Let us denote $A_0 = k$, $S_0 = qk$ and put

$$S_n = \sum_{l=0}^{q-1} A_l \exp\left(2\pi i \frac{ln}{q}\right), \quad n = 1, 2, \ldots, q-1.$$

Evidently $S_n = 0$ for $n = 1, 2, \ldots, q-1$. It follows that for $v \not\equiv 0 \bmod q$

$$T = \sum_{n=0}^{q-1} S_n \exp\left(-2\pi i \frac{vn}{q}\right) = S_0 = qk.$$

On the other hand, inverting the order of summations, we obtain

$$T = \sum_{l=0}^{q-1} \sum_{n=0}^{q-1} A_l \exp\left(2\pi i \frac{(l-v)n}{q}\right) = qA_v.$$

Thus it follows $A_v = k$ for $v = 1, 2, \ldots, q-1$, which was to be proved.

Lemma 6. *If $b_0, b_1, \ldots, b_{N-1}$ is a difference basis of order k modulo q, and $\mu(E)$ denotes the measure of equidistribution of the set $E = E_q(b_0, b_1, \ldots, b_{N-1})$, we have*

$$\mu(E) = 1 - \frac{N-k}{N^2}. \tag{14}$$

Lemma 6 follows from the proof (not the statement) of Lemma 4 combined with Lemma 5.

Lemma 7. *Let E denote a measurable set, $\bar{E}$ the set complementary to E. We have*

$$1-\mu(\bar{E})=\frac{1-\mu(E)}{\left(\frac{1}{|E|}-1\right)^2}. \tag{15}$$

Proof. Evidently

$$m(\bar{E})=\min\int_0^1 (1-F(x))\,(1-F(x+t))\,dx=1-2|E|+m(E) \tag{16}$$

and thus Lemma 7 follows.

Lemma 8. *If $\alpha(x)$ is integrable in $(0, 1)$, $\beta(x)$ bounded and integrable in the same interval and periodic with period 1, we have*

$$\lim_{n\to\infty}\int_0^1 \alpha(x)\,\beta(nx)\,dx=\int_0^1 \alpha(x)\,dx\int_0^1 \beta(x)\,dx. \tag{17}$$

This lemma is well known[8]).

Lemma 9. *Let E_1 and E_2 denote two sets having positive measures $|E_1|$ and $|E_2|$, characteristic functions $F_1(x)$ and $F_2(x)$, distance functions $G_1(x)$ and $G_2(x)$, respectively, and let the minima of the distance functions be denoted by $m(E_1)$ and $m(E_2)$ respectively. Let us define the set $E^{(n)}$ by its characteristic function being $F^{(n)}(x)=F_1(x)\,F_2(nx)$ $(n=1, 2, \ldots)$. It follows*

$$\lim_{n\to\infty}|E^{(n)}|=|E_1|\,|E_2| \tag{18}$$

and

$$\overline{\lim_{n\to\infty}}\; m(E^{(n)})\geqq m(E_1)\,m(E_2), \tag{19}$$

where $m(E^{(n)})$ denotes the minimal value of the distance function $G^{(n)}(t)$ of $E^{(n)}$.

Proof. (18) follows clearly from Lemma 8. As regards to (19), let us suppose the contrary. Thus we suppose that there exists an infinite sequence of integers n_k $(k=1, 2, \ldots)$, and a corresponding sequence of real numbers t_{n_k} $(0\leqq t_{n_k}<1)$, for which

$$G^{(n_k)}(t_{n_k})<m(E_1)\,m(E_2)-\varepsilon$$

holds, for some fixed $\varepsilon>0$. Let us denote by τ_{n_k} the fractional part of $n_k t_{n_k}$. Clearly we may choose an infinite subsequence ν_k $(k=1, 2, \ldots)$ of the sequence n_k, such that if $k\to\infty$, t_{ν_k} and τ_{ν_k} tend to limits t^* and τ^*,

---

8) This lemma has been proved for some special cases by L. Fejér, Lebesguesche Konstanten und divergente Fourierreihen, *Journal für reine und angewandte Math.*, **138** (1910), pp. 27–28. In the general form the lemma has been proved by A. Zygmund, l. c. (1), p. 173, § 8. 34.

respectively. Now, putting

$$G_k(t^*, \tau^*) = \int_0^1 F_1(x)\, F_1(x+t^*)\, F_2(\nu_k x)\, F_2(\nu_k x + \tau^*)\, dx,$$

we have

$$|G^{(\nu_k)}(t_{\nu_k}) - G_k(t^*, \tau^*)| \leqq \int_0^1 |F_1(x+t_{\nu_k}) - F_1(x+t^*)|\, dx + \\ + \int_0^1 |F_2(y+\tau_{\nu_k}) - F_2(y+\tau^*)|\, dy$$

and thus, applying again the theorem by which we have proved the continuity of $G(t)$ (see [1])), we obtain

$$\lim_{k\to\infty} [G^{(\nu_k)}(t_{\nu_k}) - G_k(t^*, \tau^*)] = 0. \tag{21}$$

Applying Lemma 8 again, we obtain

$$\lim_{k\to\infty} G_k(t^*, \tau^*) = G_1(t^*)\, G_2(\tau^*) \geqq m(E_1)\, m(E_2) \tag{22}$$

and thus owing to (21) it follows

$$\lim_{k\to\infty} G^{(\nu_k)}(t_{\nu_k}) \geqq m(E_1)\, m(E_2). \tag{23}$$

But this clearly contradicts (20) and thus (19) is proved.

Lemma 10. *If the characteristic functions $F_1(x)$ and $F_2(x)$ of the measurable sets E_1 and E_2 are equal except on a set of measure $\frac{\delta}{2}$ $(0<\delta<1)$, we have $|m(E_1) - m(E_2)| < \delta$.*

Lemma 10 follows simply by remarking that $F_1(x)\, F_1(x+t)$ and $F_2(x)\, F_2(x+t)$ are equal if neither x nor $x+t$ does belong to the exceptional set, i. e. except for a set the measure of which does not exceed δ, and thus $|G_1(t) - G_2(t)| < \delta$ for any t. Let $M(\alpha)$ denote the least upper bound of $\mu(E)$ for all sets E for which $|E| = \alpha$ $(0<\alpha<1)$. We prove

Lemma 11.

If $M(\alpha) = 1$ and $M(\beta) = 1$, we have $M(\alpha\beta) = 1$.

Proof. According to the suppositions of our Lemma, for any $\varepsilon > 0$ there exist sets E_1 and E_2 with $|E_1| = \alpha$, $|E_2| = \beta$, $\mu(E_1) > 1 - \frac{\varepsilon}{4}$, $\mu(E_2) > 1 - \frac{\varepsilon}{4}$. Let us define the sequence of sets $E^{(n)}$ as in Lemma 9, by virtue of which we have $\lim_{n\to\infty} |E^{(n)}| = \alpha\beta$ and

$$\lim_{n\to\infty} m(E^{(n)}) \geqq \alpha\beta \left(1 - \frac{\varepsilon}{4}\right)^2.$$

Thus if we choose n sufficiently large, both inequalities

$$\left||E^{(n)}|-\alpha\beta\right|<\frac{\varepsilon\alpha\beta}{4} \quad \text{and} \quad m(E^{(n)})\geqq\alpha\beta\left(1-\frac{\varepsilon}{2}\right)$$

will be satisfied. According to $|E^{(n)}|-\alpha\beta<0$ or $|E^{(n)}|-\alpha\beta>0$ we may add or take away from $E^{(n)}$ a set of measure not exceeding $\frac{\varepsilon\alpha\beta}{4}$ so as to obtain a set $\mathcal{E}^{(n)}$ having its measure equal to $\alpha\beta$. The characteristic function of the set $\mathcal{E}^{(n)}$ does not differ from that of $E^{(n)}$ but on a set the measure of which does not exceed $\frac{\varepsilon\alpha\beta}{4}$. Thus, according to Lemma 10, we have

$$m(\mathcal{E}^{(n)})\geqq m(E^{(n)})-\frac{\varepsilon\alpha\beta}{2}\geqq\alpha\beta(1-\varepsilon).$$

As $\varepsilon>0$ may be chosen arbitrarily, this proves Lemma 11.

Lemma 12. *If* $\lim_{n\to\infty}\alpha_n=\alpha$ $(0<\alpha_n<1,\ 0<\alpha<1)$ *and* $M(\alpha_n)=1$ *for* $n=1,2,\ldots$, *then we have* $M(\alpha)=1$.

Proof. For any $\varepsilon>0$, we choose n sufficiently large so as to obtain

$$\left|\frac{\alpha_n}{\alpha}-1\right|<\frac{\varepsilon}{4}.$$

According to our suppositions, there exists a set E_n for which $|E_n|=\alpha_n$ and $\mu(E_n)\geqq 1-\frac{\varepsilon}{4}$. We add to or take away from E_n a set of measure not exceeding $\frac{\alpha\varepsilon}{4}$ so as to obtain a set $\mathcal{E}_n$ of measure α. We have, using Lemma 10,

$$m(\mathcal{E}_n)\geqq m(E_n)-\frac{\alpha\varepsilon}{2}\geqq\alpha_n\left(1-\frac{\varepsilon}{4}\right)-\frac{\alpha\varepsilon}{2}\geqq\alpha(1-\varepsilon)$$

which proves Lemma 12.

Lemma 13. *Every real number* α $(0<\alpha<1)$ *can be represented as a finite or infinite product of the form*

$$\alpha=\prod_{k=1}^{\infty}\left(1-\frac{1}{2^{n_k}}\right) \qquad (1\leqq n_k\leqq n_{k+1}). \tag{24}$$

Proof. Let us suppose that α is not a rational number which is equal to the product of a finite number of factors of the form $\left(1-\frac{1}{2^{n_k}}\right)$. Let us choose $n_1\geqq 1$ so that we have

$$1-\frac{1}{2^{n_1-1}}<\alpha<1-\frac{1}{2^{n_1}}; \tag{25}$$

further, if $n_1, n_2, \ldots, n_{k-1}$ are already found, we choose n_{k+1} so as to obtain

$$1-\frac{1}{2^{n_k-1}}<\frac{\alpha}{\prod_{j=1}^{k-1}\left(1-\frac{1}{2^{n_j}}\right)}<1-\frac{1}{2^{n_k}}. \tag{26}$$

Dividing (26) by $1-\frac{1}{2^{n_k}}$ and applying again (26) with $k+1$ instead of k, we obtain

$$1-\frac{1}{2^{n_k-1}}<\frac{\alpha}{\prod_{j=1}^{k}\left(1-\frac{1}{2^{n_j}}\right)}<1-\frac{1}{2^{n_{k+1}}}. \tag{27}$$

It follows from (27) that $n_{k+1}\geqq n_k$. Thus the sequence n_k, which is uniquely determined according to the above construction is non-decreasing. It is easy to see, that $n_k\to\infty$. As a matter of fact, in the opposite case n_k would be constant from some index k_0 onwards. But it would follow from the construction that in this case we should have

$$\alpha\leqq\prod_{j=1}^{k_0-1}\left(1-\frac{1}{2^{n_j}}\right)\left(1-\frac{1}{2^{n_{k_0}}}\right)^N \tag{28}$$

for any N, i. e. we should have $\alpha=0$, contrary to our hypothesis. Thus $n_k\to\infty$, and it follows from (26) that

$$\lim_{k\to\infty}\frac{\alpha}{\prod_{j=1}^{k-1}\left(1-\frac{1}{2^{n_j}}\right)}=1$$

which proves our lemma.

§. 2. Application of the theorem of Lagrange.

The theorem of LAGRANGE is question is the following: Let p denote a prime number of the form $4n+3$. Let $r_1, r_2, \ldots, r_\nu$ $\left(\nu=\frac{p-1}{2}\right)$ denote a complete system of quadratic residues mod p. Let d denote any integer, $d\not\equiv 0 \bmod p$. Then there are $\frac{p-3}{4}$ quadratic residues in the sequence r_j+d $(j=1,2,\ldots,\nu)$. According to the terminology introduced in Lemma 5, this theorem can be stated also by saying that the system of quadratic residues to a prime modulus $p\equiv 3 \bmod 4$ is a difference basis of order $\frac{p-3}{4}$ modulo p. This theorem follows easily from Lemma 5 and from the well known formula for Gaussian sums:

$$\sum_{y=0}^{p-1}\exp\left(\frac{2\pi i y^2}{p}\right)=i\sqrt{p} \tag{29}$$

for $p \equiv 3 \bmod 4$. As every class of quadratic residues is represented twice among the squares y^2 $(1 \leqq y \leqq p-1)$, it follows from (29) that

$$\sum_{j=1}^{\nu} \exp\left(\frac{2\pi i r_j}{p}\right) = \frac{i\sqrt{p}-1}{2}. \tag{30}$$

It follows from (30), using $\sum_{k=0}^{p-1} \exp \frac{2\pi i k}{p} = 0$, that if $s_1, s_2, \ldots, s_\nu$ denote a complete set of quadratic non-residues mod p, we have

$$\sum_{j=1}^{\nu} \exp\left(\frac{2\pi i s_j}{p}\right) = \frac{-i\sqrt{p}-1}{2}. \tag{31}$$

Now the sequence nr_j $(j = 1, 2, \ldots, \nu)$ is congruent to the sequence of residues or to the sequence of non-residues, according to the quadratic character of n. Thus it follows from (30) and (31) that for any $n \not\equiv 0 \bmod p$ we have

$$\left|\sum_{j=1}^{\nu} \exp\left(\frac{2\pi i n r_j}{p}\right)\right|^2 = \frac{p+1}{4}. \tag{32}$$

Thus we can apply Lemma 5, and obtain that the differences $r_i - r_j$, $i \neq j$ represent every class of residues mod p exactly $\frac{p-1}{2} - \frac{p+1}{4} = \frac{p-3}{4}$ times, which is equivalent to the theorem of LAGRANGE stated above.

Now everything is ready to prove

Theorem 1. *The least upper bound $M(\alpha)$ of the measure of equidistribution $\mu(E)$ of measurable sets E having the measure $|E| = \alpha$ is identically equal to 1 for $0 < \alpha \leqq 1$.*

Proof of Theorem 1. Let p denote a prime, $p \equiv 3 \bmod 4$, and let $r_1, r_2, \ldots, r_\nu$ $\left(\nu = \frac{p-1}{2}\right)$ denote a complete system of quadratic residues mod p. Let us define the set $E_p = E_p(r_1, r_2, \ldots, r_\nu)$ as in § 1. It follows from Lemma 6 that

$$\mu(E_p) = 1 - \frac{p+1}{(p-1)^2}. \tag{33}$$

Let $\mathcal{E}_p$ denote a set obtained by adding to E_p any interval of length $\frac{1}{2p}$. As $|E_p| = \frac{p-1}{2p}$, we have $|\mathcal{E}_p| = \frac{1}{2}$ and it follows from (33) that

$$\mu(\mathcal{E}_p) \geqq 1 - \frac{3}{p}. \tag{34}$$

Since there are an infinity of primes of the form $4n+3$, it follows that $M\left(\frac{1}{2}\right) = 1$. Applying Lemma 11, we obtain $M\left(\frac{1}{2^k}\right) = 1$ for $k = 1, 2, \ldots,$

further, by Lemma 7, $M\left(1-\frac{1}{2^k}\right)=1$, $(k=1,2,\ldots)$. Applying Lemma 11 again, we obtain that $M(\alpha)=1$ if α is a finite product of the form (24). Thus it follows, using Lemma 12 and regarding also Lemma 13, that $M(\alpha)=1$ for all α, $0<\alpha\leqq 1$. Thus Theorem 1 is proved[9]).

§. 3. The measure of k-fold equidistribution.

Let $E, E_t, |E|$ and $F(x)$ have the meaning as in the introduction. Let $G(t_1, t_2, \ldots, t_k)$ denote the measure of the set of points common to $E, E_{t_1}, E_{t_2}, \ldots, E_{t_k}$. We have evidently

$$G(t_1, t_2, \ldots, t_k)=\int_0^1 F(x)\,F(x+t_1)\,F(x+t_2)\ldots F(x+t_k)\,dx. \tag{35}$$

It is easy to see that $G(t_1, t_2, \ldots, t_k)$ is a continuous function of its variables. The minimal value of $G(t_1, t_2, \ldots, t_k)$ shall be denoted by $m_k(E)$. Owing to

$$\int_0^1\int_0^1\ldots\int_0^1 G(t_1, t_2, \ldots, t_k)\,dt_1\,dt_2\ldots dt_k=|E|^{k+1}, \tag{36}$$

we have

$$0\leqq m_k(E)\leqq |E|^{k+1}. \tag{37}$$

The measure of k-fold equidistribution of the set E shall be defined by

$$\mu_k(E)=\frac{m_k(E)}{|E|^{k+1}}. \tag{38}$$

Thus we have, owing to (37), $0\leqq\mu_k(E)\leqq 1$. The least upper bound of $\mu_k(E)$ for all measurable sets E with $|E|=\alpha$ will be denoted by $M_k(\alpha)$. It seems probable that $M_k(\alpha)=1$ identically in α for any k. In what follows we shall prove however only the following

Theorem 2.

$$\overline{\lim_{k\to\infty}}\ M_k(\alpha)\geqq\frac{1}{4^{k+1}}.$$

The most surprising consequence of Theorem 1 is perhaps that there exist measurable sets with arbitrary small positive measure with the property, that if the set is "rotated" in the sense mentioned in the introduction, the set of points, which are common to the rotating set and to the original set, is never void, indeed, its measure exceeds always a fixed number during the rotation. Though Theorem 2 is relatively

---

[9]) Mr. P. Ungár, to whom I communicated at an earlier stage of my investigations some of my results, found independently a proof of Theorem 1, running essentially on the same lines.

much weaker than Theorem 1, and is not a "best possible" result, nevertheless it contains the generalization of that interpretation of Theorem 1 which has been emphasised just now.

The proof of Theorem 3 will be based on the following generalization of a theorem of THUE:

Lemma 14. *If p is a prime, k a positive integer, further the positive integers $e_1, e_2, \ldots, e_k$, f satisfy*

$$e_1 . e_2 \ldots e_k . f > p^k, \tag{39}$$

then for any k-tuple of integers $(r_1, r_2, \ldots, r_k)$ there can be found integers $x_1, x_2, \ldots, x_k$ and y for which $1 \leqq y < f$, $|x_i| < e_i$ $(i = 1, 2, \ldots, k)$ and $r_i \equiv \frac{x_i}{y} \mod p$ $(i = 1, 2, \ldots, k)$ are valid.

Proof. Let us consider all k-tuples of integers of the form $(yr_i + x_i)$, $i = 1, 2, \ldots, k$, where $1 \leqq x_i \leqq e_i$ $(i = 1, 2, \ldots, k)$ and $1 \leqq y \leqq f$. The number of such k-tuples of integers being $e_1 e_2 \ldots e_k f$, as there are only p^k k-tuples which are different mod p, owing to (39), there must be at least two k-tuples of the form considered which are congruent mod p. If we denote the two congruent k-tuples by $(yr_i + x_i)$ and $(\eta r_i + \xi_i)$, $i = 1, 2, \ldots, k$, we have

$$yr_i + x_i \equiv \eta r_i + \xi_i \mod p, \quad i = 1, 2, \ldots, k.$$

From $y \equiv \eta \mod p$ it would follow $x_i \equiv \xi_i \mod p$ for all $i = 1, 2, \ldots, k$, thus we have $y \not\equiv \eta \mod p$, and it follows

$$r_i \equiv \frac{\pm |\xi_i - x_i|}{|y - \eta|} \mod p \qquad (i = 1, 2, \ldots, k).$$

As $0 \leqq |\xi_i - x_i| < e_i$ $(i = 1, 2, \ldots, k)$ and $1 \leqq |y - \eta| < f$, our Lemma is proved.

Now we prove the following

Lemma 15. *If p is a prime, k a positive integer, and $Q = \left[p^{\frac{k}{k+1}}\right]$ ($[x]$ denotes the integral part of x), a set of $2Q$ integers $c_1, c_2, \ldots, c_{2Q}$ can be given, having the property that for any k-tuple of integers $(b_1, b_2, \ldots, b_k)$, elements $c_{i_1}, c_{i_2}, \ldots, c_{i_k}$, c_j of the given set can be chosen so as to obtain*

$$b_r \equiv c_{i_r} - c_j \mod (p-1) \quad \text{for} \quad r = 1, 2, \ldots, k.$$

Proof. Putting $e_i = f = \left[p^{\frac{k}{k+1}}\right] + 1 = Q + 1$ $(i = 1, 2, \ldots, k)$, condition (39) of lemma 14 is evidently satisfied. Let g denote a primitive root mod p and let $\operatorname{ind} x$ denote the index of the residue class x with respect to g. It is easy to see that if $c_i = \operatorname{ind} i$, $c_{Q+i} = \operatorname{ind}(-i)$ $(i = 1, 2, \ldots, Q)$, the sequence c_i $(1 \leqq i \leqq 2Q)$ has the required properties.

Let us now define the set E, consisting of the intervals:

$$\frac{c_r-1}{p-1}\leqq x\leqq\frac{c_r+1}{p-1}, \tag{40}$$

where the c_r $(r=1,2,\ldots,2Q)$ are the elements of the set of integers of Lemma 15. Let $F(x)$ denote the characteristic function of the set E, and let $G(t_1,t_2,\ldots,t_k)$ be defined by (35). If $(t_1,t_2,\ldots,t_k)$ is an arbitrary k-tuple of real numbers, $0\leqq t_r<1$, we put

$$t_r=\frac{b_r+\vartheta_r}{p-1} \qquad (r=1,2,\ldots,k),$$

where b_r denotes the integer which is nearest to $(p-1)t_r$, and thus we have

$$|\vartheta_r|\leqq\frac{1}{2} \qquad (r=1,2,\ldots,k).$$

According to Lemma 15, we can choose $c_{i_1},c_{i_2},\ldots,c_{i_k},c_j$ so that

$$b_r\equiv c_{i_r}-c_j \mod(p-1) \quad \text{for } r=1,2,\ldots,k.$$

It follows according to (40) that if

$$\frac{c_j-\frac{1}{2}}{p-1}\leqq x\leqq\frac{c_j+\frac{1}{2}}{p-1},$$

we have

$$\frac{c_{i_r}-1}{p-1}\leqq x+t_r\leqq\frac{c_{i_r}+1}{p-1} \qquad \text{for } r=1,2,\ldots,k.$$

Thus

$$F(x+t_r)=1 \quad \text{for } r=1,2,\ldots,k \quad \text{if} \quad \frac{c_j-\frac{1}{2}}{p-1}\leqq x\leqq\frac{c_j+\frac{1}{2}}{p-1}.$$

It follows from (35) that

$$m_k(E)\geqq\frac{1}{p-1}. \tag{41}$$

Owing to $|E|=\dfrac{4Q}{p-1}$ $\left(Q=\left[p^{\frac{k}{k+1}}\right]\right)$, we obtain

$$\mu_k(E)\geqq\frac{\left(1-\frac{1}{p}\right)^k}{4^{k+1}}. \tag{42}$$

If any fixed $\varepsilon>0$ is given, we can choose p sufficiently large so as to obtain

$$|E|<\varepsilon \quad \text{and} \quad \mu_k(E)>\frac{1-\varepsilon}{4^{k+1}}. \tag{43}$$

Thus Theorem 2 is proved.

§. 4. Some remarks on the sequences of Singer.

We have seen in §. 1 that the construction of highly equidistributed sets is closely connected with the number-theoretical problem of constructing difference bases, i. e. finite sequences of integers, the differences of which represent every class of residues to a given modulus q exactly k times, k being the order of the difference basis. In this direction interesting results have been obtained by I. SINGER (l. c. [6])) who constructed difference bases of order 1 for any modulus q of the form $q=p^{2m}+p^m+1$, p prime. Let a_j $(j=0, 1, \ldots, p)$ denote such a sequence of SINGER; we may suppose evidently

$$0 \leqq a_0 < a_1 < \ldots < a_p < q.$$

It follows that for any k $(1 \leqq k < q)$ either k or $k-q$ can be represented in the form $a_i - a_j$, and we may ask which subset of $1, 2, \ldots, q-1$ is represented "actually", i. e. for which k we have $k=a_i-a_j$. This problem, in a somewhat different from, has been raised by L. RÉDEI and is discussed in a joint paper of L. RÉDEI and the author[10]) where the following theorem is proved: If η^* denotes the minimal number of terms of a finite sequence of integers with the property that their differences represent every number $1, 2, \ldots, n$, then

$$\lim_{n\to\infty} \frac{n^*}{\sqrt{n}} = \gamma \tag{44}$$

exists, further we have[11])

$$\sqrt{2+\frac{4}{3\pi}} \leqq \frac{n^*}{\sqrt{n}} \leqq \sqrt{\frac{8}{3}}. \tag{45}$$

Now these problems are also connected with the theory of equidistribution of point sets. To establish this connection, we have to define the "asymmetric distance function" $g(t)$ of a set E as follows:

Let $f(x)$ denote the characteristic function of the set E if x is contained in the interval $(0, 1)$, and let us define $f(x)=0$ for x outside of $(0, 1)$. We put

$$g(t) = \int_0^1 f(x) f(x+t)\, dx \qquad (-1 \leqq t \leqq +1). \tag{46}$$

---

[10]) To be published in the *Mat. Sbornik.*

[11]) As BÉLA SZ.-NAGY kindly remarked, the lower estimation in (45) can be improved, by some numerical refinement, by approximately 0,01. A similar remark applies to (49). P. ERDŐS and I. S. GÁL proved by some modification of the original proof that (44) and (45) are valid also if the sequence of integers in question is restricted by the condition that it is contained in the sequence $1, 2, \ldots, n$; cf. *Proceedings Koninklijke Nederlandsche Akademie van Wetenschappen*, **51** (1948), pp. 1155–1159.

It is easy to see that $g(t)$ is an even continuous function, further that $g(0)=E$, $g(1)=0$, and we have

$$\int_0^1 g(t)\,dt=\frac{|E|^2}{2}. \tag{47}$$

We obtain further by some simple calculations that

$$\int_0^1 g(t)\cos\lambda t\,dt=\frac{1}{2}\left|\int_0^1 f(x)\exp(i\lambda x)\,dx\right|^2, \tag{48}$$

i. e. that the FOURIER cosine transform of $g(t)$ is non-negative. This is the idea underlying the proof of the following property of the sequences of SINGER:

Theorem 3. *Let us denote* $P=p^m$ (p *prime*), $q=P^2+P+1$ *and* $k=\frac{q-1}{2}$. *If* $0\leqq a_0<a_1<\ldots<a_p<q$ *denotes a* SINGER *sequence, and if* $1\leqq A_1<A_2<\ldots<A_k$ *denote the numbers which are representable in the form* a_i-a_j *with* $i>j$, *further if* $A_k=k+D$ (*i. e.* D *denotes how many numbers are missing from the sequence* $1,2,\ldots,A_k$) *then we have*

$$D>\frac{P^2+1}{3\pi-2}-\frac{P}{2}. \tag{49}$$

Proof. We have

$$\left|\sum_{j=0}^{p}\exp(2\pi i a_j t)\right|^2=P+1+2\sum_{n=1}^{k}\cos 2\pi A_n t= \tag{50}$$

$$=P+\frac{\sin(2A_k+1)\frac{t}{2}}{\sin\frac{t}{2}}-2\sum_{v=1}^{D}\cos 2\pi B_v t,$$

where B_v ($v=1,2,\ldots,D$) denote the numbers $<A_k$ which are not contained in the sequence A_j. It follows from (50) that

$$0\leqq P+\frac{\sin(2A_k+1)\frac{t}{2}}{\sin\frac{t}{2}}+2D \tag{51}$$

for all values of t. Let us choose $t=\frac{3\pi}{2A_k+1}$, using $\sin x<x$ for $x>0$. We obtain

$$2D\geqq\frac{2(2A_k+1)}{3\pi}-P, \tag{52}$$

from which Theorem 3 follows by simple calculation.

It may be remarked that though (49) is not a best possible estimate, it gives a rather good estimation for small values of P. Thus the

set A_j coincides with the set $1, 2, \ldots, k$ only for $P=2$ and $P=3$ (the corresponding SINGER sequences are: $0, 1, 3$ for $P=2$ and $0, 1, 4, 6$ for $P=4$), further (49) asserts that for $P=4$ there must be at least one "gap" in the sequence A_j, and really there is exactly one "gap" if we consider the SINGER sequence $0, 2, 7, 8, 11$. For $P=5$, owing to (49), there must be at least two numbers missing from the sequence A_j, and there are really two gaps if we take the SINGER sequence $0, 1, 4, 10, 12, 17$, etc.

Some further progress could be obtained regarding the problems considered in the present paper if some more difference bases could be constructed. A necessary and sufficient condition however for the existence of a difference basis of order k modulo q, for given k and q, is not known.

We considered only sets E lying in the interval $(0, 1)$, but it is clear that the situation is the same for any bounded linear set. The problem of unbounded linear sets however is somewhat different, as it is shown by the remark, that in this case the symmetric and asymmetric distance functions $G(t)$ and $g(t)$ coincide.

My most sincere thanks are due to P. ERDŐS and L. RÉDEI for their valuable remarks.

(Received August 5, 1948.)