HOUSTON JOURNAL OF MATHEMATICS, Volume 2, No. 2, 1976.

PROBABILISTIC METHODS IN GROUP THEORY II

P. Erdös and R. R.Hall

Dedicated to the memory of A. Rényi.

Introduction. Let (G,+) be a finite Abelian group of order n, and suppose we choose k arbitrary elements $g_1, g_2, \ldots, g_k$ of G. Let us consider the 2^k sums $\epsilon_1 g_1 + \epsilon_2 g_2 + \ldots + \epsilon_k g_k$ where each $\epsilon_i = 0$ or 1. Two interesting questions present themselves: can every $g \in G$ be represented in the form $g = \epsilon_1 g_1 + \ldots + \epsilon_k g_k$, and if so, does each g have about the same number of representations?

Clearly for a particular set of elements $g_1, g_2, \ldots, g_k$, to answer these questions we should have to know about the structure of G: for example the elements $g_1, g_2, \ldots, g_k$ may all belong to a subgroup of G. So we ask instead, what can we expect to happen if we choose $g_1, g_2, \ldots, g_k$ at random, or, put another way, what can be said about these questions for almost all (that is, all but $o(n^k)$) of the possible choices of $g_1, g_2, \ldots, g_k$?

These probabilistic questions were raised by Erdös and Rényi [2]. Surprisingly, their answers depend very little on the structure of G; the fine detail does depend on the group structure as was pointed out by R. J. Miech [5]. If every element of G is of order 2, $\epsilon_1 g_1 + \epsilon_2 g_2 + \ldots + \epsilon_k g_k$ always generates a subgroup of G, and each element receives the same number of representations. This can be seen by viewing G as an appropriate vector space.

The only obviously necessary condition for an affirmative answer to the first question, whether every g can be represented, is $2^k \geqslant n$. Erdös and Rényi proved that provided

$$k\log 2 \geqslant \log n + 2\log\frac{1}{\delta} + \log\left(\frac{\log n}{\log 2}\right) + 5\log 2,$$

then for all but at most δn^k choices of $g_1, g_2, \ldots, g_k$ every $g \in G$ may be represented in the required form. This is nearly best possible, indeed it may be that without any conditions on the structure of G, it cannot be substantially improved. We hope to study this question in a later paper.

In this paper we consider the second question, concerning the number of representations. Our result is as follows.

THEOREM. *Let $R(g)$ denote the number of representations of g in the form $g = \epsilon_1 g_1 + \epsilon_2 g_2 + \ldots + \epsilon_k g_k$, where each $\epsilon_i = 0$ or 1. Let η be a fixed positive number. Then for almost all choices of the elements $g_1, g_2, \ldots, g_k$ we have*

$$(1 - \eta)2^k/n < R(g) < (1 + \eta)2^k/n$$

for every $g \in G$, provided

$$k \geqslant \frac{\log n}{\log 2}\left(1 + 0\left(\frac{\log\log\log n}{\log\log n}\right)\right).$$

The constant implied by the 0-notation depends only on η. Moreover, the result holds if $\eta \to 0$ as $n \to \infty$, provided $\log 1/\eta = 0(\log n/\log\log n)$.

This result is sharp except for the 0-terms, and these could be improved if the estimate for max $R(g)$ in Lemma 3 were reduced. We hope to return to this question in the future.

Erdös and Rényi [2], Miech [5], Hall [3] and Hall and Sudbery [4] have proved partial results in this direction, also Bognár [1] and Wild [6] obtained results when ϵ_i may be chosen from some fixed set of integers other than $\{0,1\}$. Erdös and Rényi proved that it is sufficient that $k\log 2 \geqslant 2\log n + 2\log 1/\eta + \phi(n)$ where $\phi(n) \to \infty$ arbitrarily slowly as $n \to \infty$, and the subsequent work aimed at reducing the factor 2 mulitplying logn on the right. These improvements all depended on conditions on the group structure, and Erdös and Rényi conjectured that without such conditions, the factor 2 could not be reduced.

We should like to acknowledge the kind help of Professor G. L. Watson, who provided the important Lemma 1 below.

NOTATION. The language of probability is appropriate in our arguments. We write prob(...) for the probability of the event in brackets; as usual prob(A|B) means the probability of the event A, given that the event B occurs. E(...) denotes the expectation of the random variable in brackets. $E_0E_1E_2\ldots$ means the joint occurrence of the events $E_0, E_1, E_2, \ldots$.

LEMMA 1. *Let G be a finite Abelian group of order n, and suppose we are given N distinct equations*

$$\epsilon_{t,1}g_1 + \epsilon_{t,2}g_2 + \ldots + \epsilon_{t,m}g_m = 0 \quad (1 \leqslant t \leqslant N)$$

where every $\epsilon_{t,i} = 0$ or 1, $N \leqslant 2^m$. Then the number of choices of the elements $g_1, g_2, \ldots, g_m$ to satisfy all the equations simultaneously does not exceed n^{m-s}, where

$s = (\log N)/(\log 2)$.

PROOF. Let r be the unique integer such that $2^{m-r} < N \leq 2^{m-r+1}$. Select any r integers k_j, $1 \leq k_1 < k_2 < \ldots < k_r \leq m$. Since there are only 2^{m-r} choices of the coefficients $\{\epsilon_{t,i}, 1 \leq i \leq m$, i not equal to any $k_j\}$, and N equations, we can find two equations, say the t-th and u-th such that $\epsilon_{t,i} = \epsilon_{u,i}$ for every i other than the k_j's. Subtracting, we obtain an equation

(1) $v_1 g_{k_1} + v_2 g_{k_2} + \ldots + v_r g_{k_r} = 0$,

where each $v_j = 0$ or ± 1, not all zero. Now let ρ be the largest number for which there exist distinct numbers $k_1, k_2, \ldots, k_\rho$ for which no relation like (1) can be found. We have $\rho \leq r - 1$, moreover, given any other number k_0, $1 \leq k_0 \leq m$ we can deduce, from the original N equations, an equation

$$v_0 g_{k_0} + v_1 g_{k_1} + \ldots + v_\rho g_{k_\rho} = 0, \quad v_0 = \pm 1.$$

Therefore once the group elements $g_{k_1}, g_{k_2}, \ldots, g_{k_\rho}$ have been chosen, the other g_i's may be determined. Hence the equations have at most n^ρ solutions, where $\rho \leq r - 1 = [m - (\log N)/(\log 2)] \leq m - s$.

LEMMA 2. *Let $\ell = [(\log n)/(\log 2)]$, and suppose elements $g_1, g_2, \ldots, g_\ell$ are chosen randomly, and independently, from G. For each $g \in G$, let $R(g)$ denote the number of representations of g in the form $g = \epsilon_1 g_1 + \epsilon_2 g_2 + \ldots + \epsilon_\ell g_\ell$, where each $\epsilon_i = 0$ or 1. Let m be a positive integer. Then*

$$E\left(\frac{1}{n}\sum_g R^m(g)\right) \leq 2^{2^m}.$$

PROOF. Let χ denote a group character on G, so that $\chi(a+b) = \chi(a)\chi(b)$ for every $a, b \in G$. Then

$$R(g) = \frac{1}{n}\sum_\chi \bar{\chi}(g) \prod_j (1 + \chi(g_j))$$

where the product runs over $1 \leq j \leq \ell$. Hence

$$\frac{1}{n}\sum_g R^m(g) = \frac{1}{n^m}\sum_{\chi_1}{}' \ldots \sum_{\chi_m}{}' \prod_i \prod_j (1 + \chi_i(g_j)),$$

where i runs over $1 \leq i \leq m$, and Σ' denotes summation restricted by the relation $\chi_1 \chi_2 \cdots \chi_m = \chi_0$, the principal character. Therefore

$$E\left(\frac{1}{n}\sum_g R^m(g)\right) = \frac{1}{n^m}\sum_{\chi_1}{}' \ldots \sum_{\chi_m}{}' \left\{\frac{1}{n}\sum_h \prod_i (1 + \chi_i(h))\right\}^\ell,$$

the inner sum being over every group element h. But

$$\frac{1}{n}\sum_h \prod_i (1 + \chi_i(h)) = N(\chi_1, \chi_2, \ldots, \chi_m),$$

where $N(\chi_1, \chi_2, \ldots, \chi_m)$ denotes the number of distinct relations

(2) $\chi_1^{\epsilon_1}\chi_2^{\epsilon_2}\cdots\chi_m^{\epsilon_m} = \chi_0 \quad (\epsilon_i = 0 \text{ or } 1)$

existing between these characters. The characters form a group $(\hat{G},X)$ isomorphic (if we change the group operation) to $(G,+)$. Let us denote by $M_m(\hat{G},N)$ the number of choices of m characters χ_i from $\hat{G}$ to satisfy any set of exactly N relations (2). Then

(3) $E(\frac{1}{n}\sum_g R^m(g)) \leqslant \frac{1}{n^m}\sum_N M_m(\hat{G},N)N^\ell$,

summation being over the range $1 \leqslant N \leqslant 2^m$, since any set of χ's satisfy at least one relation, the empty one. We have dropped the condition $\chi_1\chi_2 \cdots \chi_m = \chi_0$ which actually implies $N \geqslant 2$, N even. For each N, there are at most $\binom{2^m}{N}$ sets of N relations, and, given such a set of relations, the number of choices of $\chi_1,\chi_2,\ldots,\chi_m$ to satisfy them does not exceed n^{m-s}, by Lemma 1, where $s = (\log N)/(\log 2)$. Hence

(4) $M_m(\hat{G},N) \leqslant \binom{2^m}{N}n^{m-s} = n^m\binom{2^m}{N}N^{-(\log n)/(\log 2)}$.

Since $N \geqslant 1$ and $\ell \leqslant (\log n)/(\log 2)$, we obtain the result stated from (3) and (4).

LEMMA 3. *Suppose elements $g_1,g_2,\ldots,g_\ell$ are chosen randomly and independently from G: ℓ and $R(g)$ are as defined in the previous lemma. Then for any fixed $A > 2$,*

$$\mathrm{Prob}(\max_g R(g) > A^{\log n/\log\log n}) \leqslant c(A)n^{-\delta(A)}$$

where $\delta(A)$ and $c(A)$ are positive numbers depending on A only.

PROOF. By Lemma 2, and Markoff's inequality, the probability in question does not exceed

$$n \cdot 2^{2^m} \cdot A^{-m\log n/\log\log n}.$$

Since $A > 2$, we can find a constant a such that $2^a < e < A^a$, and we set $m = [a\log\log n]$. The above expression tends to zero as fast as $n^{-\delta}$, where $\delta = \delta(A) = \frac{1}{2}\log(A^a/e)$.

LEMMA 4. *Suppose k elements, $g_1,g_2,\ldots,g_k$ are chosen from G randomly and independently, and $R(g)$ denotes the number of representations of g in the form $g = \epsilon_1 g_1 + \epsilon_2 g_2 + \ldots + \epsilon_k g_k$, each $\epsilon_i = 0$ or 1. Then*

$$E(\sum_g (R(g) - 2^k/n)^2) = 2^k(1 - 1/n).$$

This is equation 1.3 of Erdős and Rényi [2].

LEMMA 5. *Let H be an arbitrary but fixed sub-set of G of cardinality $|H|$. Suppose that the elements $g_1,g_2,\ldots,g_s$ are chosen randomly and independently from G, and that $N(g)$ denotes the number of choices of $\epsilon_1,\epsilon_2,\ldots,\epsilon_s$ such that $g - \epsilon_1 g_1 - \epsilon_2 g_2 -$*

... - $\epsilon_s g_s \in H$, *where each* $\epsilon_i = 0$ *or* 1. *Then*

$$E(\sum_g N^2(g) - N(g)) = n^{-1}|H|^2(4^s - 2^s).$$

PROOF. Plainly $\sum_g N(g) = 2^s|H|$. Next,

$$N(g) = \frac{1}{n}\sum_\chi \overline{\chi}(g)P(\chi,H)\prod_{i=1}^{s}(1 + \chi(g_i))$$

where

$$P(\chi,H) = \Sigma\{\chi(h): h \in H\}.$$

Therefore

$$\sum_g N^2(g) = \frac{1}{n}\sum_\chi |P(\chi,H)|^2 \prod_{i=1}^{s} |1 + \chi(g_i)|^2,$$

and

$$E(\sum_g N^2(g)) = \frac{1}{n}\sum_\chi |P(\chi,H)|^2 \{\frac{1}{n}\sum_g |1 + \chi(g)|^2\}^s.$$

But

$$\frac{1}{n}\sum_g |1 + \chi(g)|^2 = 2 \text{ if } \chi \neq \chi_0,\ 4 \text{ if } \chi = \chi_0,$$

where χ_0 is the principal character. Moreover

$$\frac{1}{n}\sum_\chi |P(\chi,H)|^2 = |H|.$$

Therefore,

$$E(\sum_g N^2(g)) = 2^s|H| + n^{-1}|H|^2(4^s - 2^s).$$

Subtracting the expectation of $\Sigma N(g)$, we obtain our result.

PROOF OF THE THEOREM. Let $\eta(\eta > 0)$ be given, and fixed. We also fix an arbitrary $A > 2$.

We begin by choosing just $\ell = [(\log n)/(\log 2)]$ elements of G. Here and in what follows we mean that the elements are chosen independently, so that repetitions can occur, and randomly: every element has an equal probability of being chosen. Let $R_0(g)$ denote the number of representations of any group element g in terms of these elements, and denote by E_0 the event

$$\max_g R_0(g) \leqslant A^{\log n/\log\log n}.$$

We now choose a further $6t + 1$ elements from G, where t is the smallest integer such that

$$2^t \geqslant A^{\log n/\log\log n}.$$

We have $k_1 = \ell + 6t + 1$ elements so far, and we denote by $R_1(g)$ the number of representations of g in terms of all of these. We call g 1-exceptional if one of the inequalities

$$(1 - \eta')\frac{2^{k_1}}{n} < R_1(g) < (1 + \eta')\frac{2^{k_1}}{n}$$

fails to hold, where

(5) $\eta' = \eta/2\log\log\log n$.

Let N_1 denote the number of 1-exceptional elements. Plainly

$$\sum_g (R_1(g) - 2^{k_1}/n)^2 \geqslant \eta'^2 4^{k_1} N_1/n^2$$

and we deduce from Lemma 4, and Markoff's inequality, that

$$\text{prob}(N_1 > n/2^{5t}) < 1/\eta'^2 2^t.$$

Let E_1 denote the event $N_1 \leqslant n/2^{5t}$. From the above, and Lemma 3,

$$\text{prob}(E_0E_1) > 1 - c(A)n^{-\delta(A)} - 1/\eta'^2 2^t.$$

Assume that E_0 and E_1 occur. Let H_1 denote the set of 1-exceptional elements, so that $|H_1| = N_1$. Moreover, if $g \in H_1$, we have

(6) $0 \leqslant R_1(g) \leqslant 2^{6t+1} A^{\log n/\log\log n}$.

We now choose a further s elements from G at random, giving a total of $\ell + 6t + 1 + s$, and we denote by $R_2(g)$ the number of representations of g in terms of all of these. Here s is the smallest integer such that

(7) $2^{s-1} \geqslant \frac{1}{\eta'} A^{\log n/\log\log n}$.

We call g 2-exceptional if one of the inequalities

$$(1 - \eta')^2 \frac{2^{k_2}}{n} < R_2(g) < (1 + \eta')^2 \frac{2^{k_2}}{n}$$

fails to hold, where $k_2 = \ell + 6t + 1 + s$. N_2 denotes the number of 2-exceptional elements.

Suppose that the s elements just chosen are $g_1, g_2, \ldots, g_s$, and that g has the property that for at most one choice of the numbers $\epsilon_1, \epsilon_2, \ldots, \epsilon_s$ (each $\epsilon_i = 0$ or 1) we have $g - \epsilon_1 g_1 - \epsilon_2 g_2 - \ldots - \epsilon_s g_s \in H_1$. Then

$$R_2(g) > (2^s - 1)(1 - \eta')2^{k_1}/n > (1 - \eta')^2 2^{k_2}/n$$

by (7). Also

$$R_2(g) < 2^s(1 + \eta')2^{k_1}/n + 2^{6t+1} A^{\log n/\log\log n}$$

by (6). We have $2^{k_1}/n > 2^{6t}$ by definition of k_1 and ℓ. Now using the definition of s given by (7), we deduce that

$$R_2(g) < (1 + \eta')^2 2^{k_2}/n.$$

Let $N_1(g)$ denote the number of choices of the numbers $\epsilon_1, \epsilon_2, \ldots, \epsilon_s$ above such that $g - \epsilon_1 g_1 - \epsilon_2 g_2 - \ldots - \epsilon_s g_s \in H_1$. We have shown that if g is 2-exceptional, we must have $N_1(g) \geqslant 2$. Hence

$$2N_2 \leqslant \sum_g (N_1^2(g) - N_1(g)).$$

Applying Lemma 5, and Markoff's inequality, we have that

$$\text{prob}(N_2 > n/2^{7t-4} | E_0E_1) < \frac{n \cdot 2^{-10t} \cdot 4^{s-2}}{2n \cdot 2^{-7t}} < \frac{1}{\eta'^2 2^{t+1}}$$

using the definitions of s and t. If E_2 denotes the event $N_2 \leqslant n/2^{7t-4}$, we have

$$\text{prob}(E_0E_1E_2) > 1 - c(A)n^{-\delta(A)} - (1 + \tfrac{1}{2})/\eta'^2 2^t.$$

Let H_2 denote the set of 2-exceptional elements. We have $|H_2| = N_2$, moreover, if $g \in H_2$ then

$$0 \leqslant R_2(g) \leqslant 2^{6t+1+s} A \log n/\log\log n.$$

We now choose the same number, s, random elements of G, so that we have $k_3 = \ell + 6t + 1 + 2s$. $R_3(g)$ denotes the number of representations of g in terms of all these, and we call g 3-exceptional if one of the inequalities

$$(1 - \eta')^3 2^{k_3}/n < R_3(g) < (1 + \eta')^3 2^{k_3}/n$$

fails to hold. Name the new elements $g_1, g_2, \ldots, g_s$ as before, and let $N_2(g)$ denote the number of choices of $\epsilon_1, \epsilon_2, \ldots, \epsilon_s$ for which $g - \epsilon_1 g_1 - \epsilon_2 g_2 - \ldots - \epsilon_s g_s \in H_2$. Assume that E_0, E_1, E_2 occur. Then we may check that g is 3-exceptional implies $N_2(g) \geqslant 2$. Let N_3 denote the number of 3-exceptional elements. Applying Lemma 5 and Markoff's inequality as before, we have that

$$\text{prob}(N_3 > n/2^{11t-13} | E_0E_1E_2) < 1/\eta'^2 2^{t+2}.$$

We continue in this way, adding s elements at a time, and assuming that the events $E_0, E_1, E_2, \ldots$ have all occurred. We call g r-exceptional if one of the inequalities

$$(1 - \eta')^r 2^{k_r}/n < R_r(g) < (1 + \eta')^r 2^{k_r}/n$$

fails to hold, where $k_r = \ell + 6t + 1 + (r - 1)s$. N_r denotes the number of r-exceptional elements, and we prove successively that

$$\text{prob}(N_r > n/2^{a_r t - b_r} | E_0E_1 \ldots E_{r-1}) < 1/\eta'^2 2^{t+r-1},$$

where a_r and b_r are determined from the recurrence formulae:-

$$a_{r+1} = 2a_r - 3, \; a_1 = 5; \; b_{r+1} = 2b_r + r + 3, \; b_1 = 0.$$

Plainly

$$a_r = 2^r + 3, \; b_r = 5 \cdot 2^{r-1} - r - 4.$$

We denote by E_r the event

$$N_r \leqslant n/2^{a_r t - b_r}$$

and we have that

$$p(E_0E_1 \ldots E_r) > 1 - c(A)n^{-\delta(A)} - 2(1 - \frac{1}{2^r})/\eta'^2 2^t.$$

We set

$$r_0 = [2\log\log\log n]$$

and calculation shows that if $n \geqslant 1000$, the event E_{r_0} implies $N_{r_0} < 1$, that is, $N_{r_0} = 0$. Hence we have

$$(8)\quad (1 - \eta')^{r_0} 2^{k_{r_0}}/n < R_{r_0}(g) < (1 + \eta')^{r_0} 2^{k_{r_0}}/n$$

for every $g \in G$, where $k_{r_0} = \ell + 6t + 1 + (r_0 - 1)s$. Let $k \geqslant k_{r_0}$. We may certainly choose k elements from G randomly and independently by choosing the first k_{r_0} of them in the manner described, and then choosing the rest, and we deduce from (8), inserting the values of r_0 and η' (given by (5)), that for every g, we have

$$(1 - \eta)2^k/n < R(g) < (1 + \eta)2^k/n$$

with probability at least

$$1 - c(A)n^{-\delta(A)} - 2/\eta'^2 2^t.$$

This tends to 1 as $n \to \infty$ for any fixed $\eta > 0$, indeed if

$$\frac{1}{\eta} < B^{\log n/\log\log n}$$

for any fixed B: for we may suppose $A > B^2$, and this makes 2^t tend to infinity sufficiently rapidly. We require that

$$k \geqslant k_{r_0} = \frac{\log n}{\log 2}\left(1 + 0\left(\frac{\log\log\log n}{\log\log n}\right)\right).$$

where the constant implied by the 0-notation depends on A and B only. This completes the proof.

REFERENCES

1. K. Bognár, *On a problem of statistical group theory*, Studia Sci. Math. Hungarica, 5(1970), 29-36.
2. P. Erdös and A. Rényi, *Probabilistic methods in group theory*, Journal d'Analyse Math., 14(1965), 127-138.
3. R. R. Hall, *On a theorem of Erdös and Rényi concerning Abelian groups*, J. London Math. Soc., (2), 5(1972), 143-153.
4. R. R. Hall and A. Sudbery, *On a conjecture of Erdös and Rényi concerning Abelian groups*, J. London Math. Soc., (2), 6(1972), 177-189.
5. R. J. Miech, *On a conjecture of Erdös and Rényi*, Illinois J. Math., 11(1967), 114-127.
6. K. Wild, *A theorem concerning products of elements of Abelian groups*, Proc. London Math. Soc., (3), 27(1973), 600-616.

Imperial College of Science and Technology
London, England

University of York
York, England

Received December 1, 1975