Acta Cryst. (1959). **12**, 824

Randomised and pseudorandomised substantialization of sign sequences.

By I. J. Good, *Admiralty Research Laboratory, Teddington, Middlesex, England.*

(*Received* 8 *May* 1959)

The problem of substantialization of sign sequences arises when we wish to determine the signs of the structure factors of a centrosymmetrical crystal by means of the trial analysis method (Woolfson, 1954). Some of the signs may arbitrarily be assumed to be positive (Lipson & Cochran, 1953), and I shall suppose that n' signs remain to be determined, and write $n=n'+1$. Woolfson was interested in the case $n=8$. He exhibited a set of 16 heptuples (or 7-sequences) such that each of the 128 possible heptuples differs from one of the sixteen in at most one sign. Good (1955) analysed the mathematical basis of the method and showed that it could be generalised to any value of n that is a power of 2. I have since discovered that the mathematics of the method is the same as that used in Hamming's error-correcting code (Shannon, 1948; Golay, 1949). This code was introduced for quite different purposes, namely for the reliability of binary communication. Much of the later work on error-correcting codes should also have application to crystallography.

The set of sixteen heptuples referred to above may be described as a 1-substantialization of the 128 possible heptuples, since one incorrect sign is permitted. It is perfectly economical in the sense that each of the 128 possible heptuples is 'represented' by exactly one member of the substantializing set. It is only for special values of r and n' that a perfectly economical r-substantialization of all n'-tuples exists. A necessary condition is that

$$1+\binom{n'}{1}+\ldots+\binom{n'}{r}$$

should be a power of 2; since this is the number of n'-tuples represented by one n'-tuple, and must therefore divide $2^{n'}$. When a perfectly economical substantialization does not exist we may have to be satisfied with a fairly economical one.

It will not always be possible to guess in advance what value of r should be used. In such circumstances it seems reasonable to start with a value of r not much less than $\frac{1}{2}n$ and to decrease r gradually until success is achieved. (The larger values of r provide less work, but less chance of success.)

An alternative method, which is logically simpler and therefore probably easier to apply, is to try out the n'-tuples in random order. If the time of generation of the n'-tuples is ignored this random method may be expected to take exactly twice as long as a perfectly economical substantialization using the largest value of r that would lead to success. Since this value of r is not known in advance, 'randomised substantialization' is rather better than this estimate suggests. (This factor of 2 is easy to demonstrate rigorously; it arises primarily from the fact that the systematic method would on the average succeed after half the possibilities had been tried.)

But it is better to run through the n'-tuples in a *pseudorandom* order, i.e. an order that looks random if we do not know the method of generation. I shall describe some pseudorandom orders such that no n'-tuple is repeated. Owing to this avoidance of repetition we gain back the factor of 2 that would be lost by the use of a random order of generation of the n'-tuples.

If we knew in advance that we had to have $r=0$, i.e. that substantialization was going to gain nothing, then we might as well run through the n'-tuples in the natural 'dictionary' order. The fact that most pairs of adjacent n'-tuples would strongly resemble one another would then not matter. But if we do not know that $r=0$ is required then this naive method could be expected to be very uneconomical.

The pseudorandom method has the further advantage over the random method that the workings can be more conveniently checked and described, and if we stop a piece of work in the middle we can easily remember where to begin again later. These advantages are familiar when using pseudorandom numbers in Monte Carlo methods of calculation, outside crystallography.

Of the various methods of generating pseudorandom binary numbers there is one that has a special advantage for the present purpose, namely that it is *exhaustive*. By this I mean that every n-tuple is reached sooner or later (without repetition), provided that we identify each n-tuple with its 'ones-complement'. (I am thinking of every positive sign as represented by a 1 and every negative sign by a 0, so that each n-tuple may be interpreted as a binary integer. The ones-complement of an n-tuple is then obtained by changing each 1 into a 0 and each 0 into a 1.) For some values of n we can achieve our aim with the help of Mersenne primes, i.e. prime numbers that are 1 less than a power of 2. (The use of Mersenne primes for the production of pseudorandom numbers was suggested by Lehmer (1951).)

The only values of p less than 500, for which 2^p-1 is a prime number are (Lehmer, 1953)

2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, and 127 .

Suppose that n is equal to one of these values of p, so that 2^n-1 is a prime number, say q. Let g be a primitive root of q, i.e. a number whose powers run through all the residues of q except 0. (For the terminology see the Appendix.) Then take, as our pseudorandom order, not $g, g^2, g^3, \ldots$, but instead $g^2, g^4, g^6, \ldots$, reduced mod q and expressed in the binary notation. I say that this will provide an exhaustive pseudorandomization of the n-tuples, if each n-tuple is identified with its ones-complement, except that the all-zero n-tuple is omitted. (It can be appended at the beginning if desired.) In order to prove this assertion it is sufficient to observe that the ones-complement of an n-tuple, regarded as a binary number, is simply minus that number modulo q, and that -1 is never a quadratic residue of a prime of the form $4m+3$, so the method of generation cannot produce an n-tuple *and* its ones-complement. (I am ignoring the trivial case $p=2$.)

Next suppose that n is 1 plus any of the above values of p. Then we may fix one of the signs of the n-tuple, select a primitive root of $2^{n'}-1$, and take *all* powers of

this primitive root modulo $2^{n'}-1$. We have thus coped with the following values of n: 2, 3, 4, 5, 6, 7, 8, 13, 14, 17, 18, 19, 20, 31, 32, ...

Other values of n may be dealt with in a slightly less elegant manner that will be exemplified by the case $n=10$. Let the first three binary digits, regarded as a binary number, run periodically through the sequence 4, 5, 6, 7, 4, 5, 6, 7, etc., and let the last seven binary digits, also regarded as a binary number, run periodically through the sequence $0, g, g^2, g^3, \ldots, 1, 0, g, g^2, g^3, \ldots$. Since 4 is prime to 2^7-1, the length of the entire period is 2^9-4, the only omitted 10-tuples, beginning with a 1, being

1001111111
1011111111
1101111111
1111111111.

These four 10-tuples may be appended if desired.

This method should be adequate to deal with all practical values of n not covered by the above list.

I do not know whether a list of primitive roots of Mersenne primes is already available. If not, it would be very easy to obtain them with the aid of an electronic computer, up to $p=31$.

For application of similar methods for non-centrosymmetrical crystals it may be of value to know more about the prime factors of numbers of the form 3^m-1, say. I do not know whether this problem has interested number theoreticians.

APPENDIX

Terminology of the theory of numbers

For the convenience of readers who are not familiar with the elementary theory of numbers I here list all the relevant terminology and other facts.

A prime number is an integer, $q(q \geq 2)$, not divisible by any other integer except 1. Two integers are said to be equal or congruent modulo or mod q and to belong to the same residue class or residue of q if they differ by a multiple of q. Each residue class of q can clearly be represented by one of the numbers $0, 1, 2, \ldots, q-1$.

According to Fermat's 'little theorem', if q is prime and a is not a multiple of q, then a^{q-1} is congruent to 1 mod q. For example, 3^6-1 is a multiple of 7.

A primitive root of a prime number q is a number g such that $g, g^2, g^3, \ldots, g^{q-1}$ runs through all residues of q except 0. Every prime number has at least one primitive root.

A quadratic residue of a prime, q, is a residue that is congruent to a square of an integer. The product of two quadratic residues is clearly a quadratic residue. The binary representation of an integer N is exemplified by $13=1101$, which means $1\cdot2^3+1\cdot2^2+0\cdot2^1+1\cdot2^0$, just as in the decimal representation 13 means $1\cdot10^1+3\cdot10^0$. Most modern electronic computers work internally with binary representations.

I am indebted to the Admiralty for permission to publish this paper.

References

GOLAY, M. J. E. (1949). *Proc. Inst. Radio Engrs.* **37**, 637.
GOOD, I. J. (1955). *Acta Cryst.* **7**, 603.
LEHMER, D. H. (1951). *Second Harvard Symp. on Large Scale Computing Machinery* (1949), p. 145. Harvard.
LEHMER, D. H. (1953). *Math. Tables and other Aids to Calculation*, **7**, 72.
LIPSON, H. & COCHRAN, W. (1953). *The Determination of Crystal Structures*, p. 205. London: Bell.
SHANNON, C. E. (1948). *Bell Syst. Tech. J.* **27**, § 17.
WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 65.

Notes and News

Announcements and other items of crystallographic interest will be published under this heading at the discretion of the Editorial Board. Copy should be sent direct to the Editor (P. P. Ewald, Polytechnic Institute of Brooklyn, 333 Jay Street, Brooklyn 1, *N.Y., U.S.A.) or to the Technical Editor (R. W. Asmussen, Chemical Laboratory B of the Technical University of Denmark, Sølvgade* 83, *Copenhagen K, Denmark)*

International Union of Crystallography

1. The Executive Committee very much regrets to announce that the Editor of *Acta Crystallographica*, Prof. P. P. Ewald, has requested to be released from his responsibilities by the end of the current year. The question of his succession was given the most thorough consideration before and at the meeting in Leningrad. The Executive Committee then decided, in accordance with Statutes 5·4 and 6·1, to appoint the present Editor of *Structure Reports*, Prof. A. J. C. Wilson, as Editor of *Acta Crystallographica*, and to appoint Dr W. B. Pearson as successor to Prof. Wilson as Editor of *Structure Reports*. Both appointments will take effect as from 1 January 1960.

2. Another important decision taken by the Executive Committee at its meeting in Leningrad was a readjustment of the prices of Volumes 1–6 of *Acta Crystallographica*, and of Volumes 9–13 of *Structure Reports*, together with the introduction of reduced personal prices for this latter publication for *bona-fide* crystallographers in countries adhering to the Union. As from 1 January 1960 the prices will be as follow:

Acta Crystallographica

Volumes 1–4:	
Regular price per volume	D.Cr. 100 (£5 or $14)
Reduced price for individuals	D.Cr. 60 (£3 or $ 9)
As from Volume 5:	
Regular price per volume	D.Cr. 180 (£9 or $25)
Reduced price for individuals	D.Cr. 100 (£5 or $14)