than r_{ij}. This effect was largest for small values of r_{ij} and amounted to about 1 Å for $r_{ij} = 0$ Å and $M = 0.6$ Å^{-2}. Fig. 2 shows the magnitude of these shifts as a function of r_{ij} for two representative values of M. Secondly, the relative maximum peak heights decrease with increasing r_{ij}, as shown in Fig. 3. Although calculations have been made for only two M values, 0.3 and 0.6 Å^{-2}, the results show that both of the above-mentioned effects decrease in importance with increasing M, that is, with smaller temperature factors. With single-crystal data, for example, a Gaussian distribution would probably give a good approximation to the peak shapes of a cylindrical Patterson projection for all values of r_{ij}.

It should be noted that at large values of r_{ij} (over 4.0 Å for $M = 0.6$ Å^{-2}), the shape given by (6) is closely approximated by a Gaussian function whose maximum is displaced to a larger value of R by an amount given by extrapolation of a curve like that shown in Fig. 2. The larger the value of M, the smaller is the shift and also the smaller the value of r_{ij} at which the approximation can be made. The maximum of the shifted Gaussian is similarly given by extrapolation of a curve such as that shown in Fig. 3.

Some idea of the appropriate value of M to use in the construction of a synthetic cylindrical Patterson projection can be obtained by examination of the origin peak of the observed Patterson to which the theoretical function is compared. Equation (6) has a simple form for $r_{ij} = 0$ and the variation of the position of the maximum with M may be easily derived. For example, a value of 0.6 Å^{-2} for M was suggested by the origin peak shape in the observed cylindrical Patterson projection for poly-γ-methyl-L-glutamate. This corresponds to a value of about 30 Å^2 for the temperature coefficient B, a not altogether unanticipated figure for fibrous substances with a relatively large degree of disorientation.

The complete calculation of $Q_{ij}(R, z, r_{ij}, z_{ij})$ for a set of interatomic vectors from a given model is obviously a tedious task, but some applications of IBM digital computing methods to the problem shorten the time required for such a calculation considerably.

References

JAMES, R. W. (1948). *Optical Principles of the Diffraction of X-rays.* London: Bell.

MACGILLAVRY, C. A. & BRUINS, E. M. (1948). *Acta Cryst.* **1**, 156.

YAKEL, H. L. & SCHATZ, P. N. To be published.

The Statistical Theory of Sign Relationships

BY M. M. WOOLFSON

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

A statistical investigation of the relation between the signs of structure factors shows that the probability of the result $s(h)s(h') = s(h+h')$ is $Q/(1+Q)$, where

$$Q = \exp(2\varepsilon^{-1}|U_h U_{h'} U_{h+h'}|) \quad \text{and} \quad \varepsilon = \sum_{j=1}^{N} n_j^2 .$$

These theoretical predictions are found to agree well with observation.

1. Introduction

Sayre (1952), Cochran (1952) and Zachariasen (1952) have shown that, for a centrosymmetrical structure, the signs of structure-factors of indices (hkl), $(h'k'l')$ and $(h+h',k+k',l+l')$ tend to be related so that the product

$$U_{hkl} U_{h'k'l'} U_{h+h',k+k',l+l'}$$

is a positive quantity. This tendency increases with the magnitude of the product of the unitary structure factors.

If the signs of some structure factors can be found unequivocally by the use of inequality relations (Harker & Kasper, 1948; Gillis, 1948; Karle & Hauptmann, 1950) then the range of known signs may be extended by assuming that the sign relationship holds when the structure factors involved are all large. It is even possible to allow for some inconsistencies which arise from occasional breakdowns of the relationship (Zachariasen, 1952).

It would obviously be useful to be able to calculate in advance the probability that the sign relationship will be obeyed, and this may be done by using the results of the theory developed in this paper.

2. The statistical theory

For a structure containing N atoms and having a centre of symmetry, the unitary structure factor $U_{h+h',k+k',l+l'}$ may be expressed as

$$U_{h+h',k+k',l+l'} = 2\sum_{j=1}^{N/2} n_j \cos 2\pi[(h+h')x_j$$
$$+(k+k')y_j+(l+l')z_j] ,$$

where

$$n_j = f_j / \sum_{j=1}^{N} f_j .$$

It will be assumed that n_j remains constant throughout reciprocal space (Harker & Kasper, 1948). For typographical brevity a one-dimensional notation will be used, and the above expression may be written

$$U_{h+h'} = 2\sum_{j=1}^{N/2} n_j \cos 2\pi(h+h')x_j . \tag{1}$$

We shall now find the form of the distribution function of values of $U_{h+h'}$ for a range of indices h and h', but imposing the condition that the product $U_h U_{h'}$ is constant.

We now write (1) as

$$U_{h+h'} = \sum_{j=1}^{N/2} \xi_j ,$$

where

$$\xi_j = 2n_j \cos 2\pi(h+h')x_j .$$

The mean value of ξ_j averaged over the complete range of h and h' will be

$$\overline{\xi_j} = 2n_j \overline{\cos 2\pi(h+h')x_j}$$
$$= 2n_j \overline{(\cos 2\pi h x_j . \cos 2\pi h' x_j - \sin 2\pi h x_j . \sin 2\pi h' x_j)} .$$

The mean value of the product of two variables is the product of their mean values, and hence

$$\overline{\sin 2\pi h x_j . \sin 2\pi h' x_j} = \overline{\sin 2\pi h x_j} \times \overline{\sin 2\pi h' x_j} = 0 .$$

To find the weighted mean value of

$$\cos 2\pi h x_j . \cos 2\pi h' x_j$$

we make use of the equation

$$U_h U_{h'} = \sum_{i=1}^{N} \sum_{j=1}^{N} n_i n_j \cos 2\pi h x_i . \cos 2\pi h' x_j .$$

$U_h U_{h'}$ may be considered as the sum of a number of variables $\cos 2\pi h x_i . \cos 2\pi h' x_j$ with corresponding weights $n_i n_j$.

Hence

$$\overline{\cos 2\pi h x_i . \cos 2\pi h' x_j} = \frac{U_h U_{h'}}{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} n_i n_j} = U_h U_{h'} .$$

The particular case $i = j$ is not different from the general case $i \neq j$ since the variables $\cos 2\pi h x_j$ and $\cos 2\pi h' x_j$ are unrelated for varying h and h'.

From this we find

$$\overline{\xi_j} = 2n_j U_h U_{h'} . \tag{2}$$

The central-limit theorem (Cramér, 1946) states that the sum of a large number, N, of independent variables ξ_j, of mean value $\overline{\xi_j}$ and mean square devia-

tion α_j^2, is normally distributed about $\sum\limits_{j=1}^{N} \overline{\xi_j}$ with mean square deviation $\sum\limits_{j=1}^{N} \alpha_j^2$.

The variables ξ_j are not strictly independent since the covariance of ξ_i and ξ_j is

$$\overline{\xi_i \xi_j} = 4n_i n_j \overline{\cos 2\pi(h+h')x_i . \cos 2\pi(h+h')x_j}$$
$$= 4n_i n_j U_h^2 U_{h'}^2 .$$

The normal range for which the sign relationships are used is that for which

$$U_h \quad \text{and} \quad U_{h'} \leqslant 0.4 .$$

For U's greater than this the range of the inequality relationships is reached and sign determination becomes a certain process.

Hence in general

$$U_h^2 U_{h'}^2 \ll 1 . \tag{3}$$

The covariance is thus very small and the variables may be taken as independent. Then the mean value of $U_{h+h'} (= \sum\limits_{j=1}^{N/2} \xi_j)$ is

$$\overline{U_{h+h'}} = \sum_{j=1}^{N/2} 2n_j U_h U_{h'} = U_h U_{h'} . \tag{4}$$

The mean square value of ξ_j averaged over all h and h' is

$$\overline{\xi_j^2} = 4n_j^2 \overline{\cos^2 2\pi(h+h')x_j}$$
$$= 4n_j^2 \overline{\tfrac{1}{2} + \tfrac{1}{2}\cos 2\pi . 2(h+h')x_j} .$$

We assume that $\overline{\cos 2\pi . 2(h+h')x_j} \ll 1$, since $U_h U_{h'}$ is small. Then $\overline{\xi_j^2} = 2n_j^2$.

The mean square deviation of ξ_j is given by

$$\alpha_j^2 = \overline{\xi_j^2} - \overline{\xi_j}^2 ,$$

which gives

$$\alpha_j^2 = 2n_j^2(1 - 2U_h^2 U_{h'}^2)$$

From (3) the bracket may be replaced by unity and

$$\alpha_j^2 = 2n_j^2 .$$

The mean square deviation of the variable $U_{h+h'}$ is thus

$$A^2 = \sum_{j=1}^{N/2} 2n_j^2 = \varepsilon ,$$

where ε is written for $\sum\limits_{j=1}^{N} n_j^2 .$

The probability of a value of $U_{h+h'}$ between U_{h+h} and $U_{h+h'} + d(U_{h+h'})$ is

$$P(U_{h+h'})d(U_{h+h'})$$
$$= (2\pi\varepsilon)^{-\frac{1}{2}} \exp\{-(2\varepsilon)^{-1}(U_{h+h'} - U_h U_{h'})^2\}d(U_{h+h'}) .$$

If we now consider the three unitary structure factors U_h, $U_{h'}$, $U_{h+h'}$ the probability that $U_{h+h'}$ will have the sign of $U_h U_{h'}$, divided by the probability of the sign being opposite, is

$$Q = \frac{P\{s(h)s(h')|U_{h+h'}|\}}{P\{-s(h)s(h')|U_{h+h'}|\}}$$

$$= \frac{\exp\left\{-(2\varepsilon)^{-1}(|U_{h+h'}|-|U_hU_{h'}|)^2\right\}}{\exp\left\{-(2\varepsilon)^{-1}(|U_{h+h'}|+|U_hU_{h'}|)^2\right\}}$$

$$= \exp\left\{2\varepsilon^{-1}|U_hU_{h'}U_{h+h'}|\right\}. \qquad (5)$$

The probability, p, that the sign relationship will hold is thus

$$p = Q/(1+Q).$$

For a structure containing N equal and resolved atoms the expression for Q becomes

$$\exp\left\{2N|U_hU_{h'}U_{h+h'}|\right\}.$$

3. Some properties of the probability relationship

(i) The effect of atomic overlap

The usual method of putting the experimental X-ray data on to an absolute scale depends on the fact that the mean reflexion intensity, $\langle I \rangle$, is given by

$$\langle I \rangle = \Sigma,$$

where

$$\Sigma = \sum_{j=1}^{N} f_j^2 \qquad \text{(Wilson, 1948)}.$$

In terms of ε this becomes

$$\langle I \rangle = \left(\sum_{j=1}^{N} f_j\right)^2 \times \varepsilon.$$

If atomic overlap occurs the effective value of Σ increases; for example, two equal atoms which completely overlap contribute twice as strongly to Σ as two similar resolved atoms. It is the effective value of Σ, Σ_e, which should be used for scaling and similarly there is an effective value of ε, ε_e, which should be used in (5).

If the root-mean-square value of the unitary structure-factors is $(\overline{U^2})^{\frac{1}{2}}$, then

$$(\overline{U^2})^{\frac{1}{2}} = \left\{\langle I \rangle \Big/ \left(\sum_{j=1}^{N} f_j\right)^2\right\}^{\frac{1}{2}} = \sqrt{(\varepsilon_e)},$$

and if we write

$$U_h/(\overline{U^2})^{\frac{1}{2}} = \gamma_h$$

then (5) becomes

$$Q = \exp\left\{2\sqrt{(\varepsilon_e)} \cdot |\gamma_h\gamma_{h'}\gamma_{h+h'}|\right\}. \qquad (6)$$

The experimental determination of the values of $|\gamma|$ is unaffected by the value of ε_e, and ε_e remains the only unknown quantity in (6). As overlap increases so does the value of ε_e, and replacing ε_e by ε gives the minimum probability of the sign relationship being true.

The fact that the efficiency of sign relationships is improved by atomic overlap has been pointed out by Cochran (1952), who derives sign relationships from the condition

$$\int_V \varrho^3 dV \qquad \text{is maximum positive.}$$

This condition will become increasingly true if regions of large ϱ are built up by atomic overlap, or by an atom of relatively great atomic number.

(ii) The effect of structural complexity

The distribution of the structure factors of a centrosymmetric structure is given by the function

$$P(F)dF = (2\pi\Sigma)^{-\frac{1}{2}} \exp\left\{-F^2/2\Sigma\right\}dF,$$

where $P(F)dF$ is the probability of a structure factor having a value between F and $F+dF$.

In terms of γ $(= F/\sqrt{\Sigma})$ we have

$$P(\gamma)d\gamma = (2\pi)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\gamma^2\right\},$$

which is independent of the number of atoms in the unit cell, or structural complexity.

A given value of the product $\gamma_h\gamma_{h'}\gamma_{h+h'}$ will thus be equally likely for all structures, and structural complexity will influence the efficiency of the sign relationship only by its effect on the value of $\sqrt{(\varepsilon_e)}$. For a structure containing N equal resolved atoms, $\sqrt{(\varepsilon_e)} = N^{-\frac{1}{2}}$. It can thus be seen that sign relationships become less powerful for more complex structures.

(iii) The effect of higher structural symmetry

A mathematical analysis similar to that of § 2 shows that the expressions for Q and p are effectively independent of the symmetry number. The degree of independence is given roughly by the extent to which

$$nU_h^2U_{h'}^2 \ll 1,$$

where n is the symmetry number.

4. Some practical illustrations of the statistical theory

(i) An empirical test of the distribution of $U_{h+h'}$

Unitary structure factors were calculated for a model structure with $n = 2$ and ten atoms in the asymmetric unit. A statistical survey was made of more than 700 values of $s(h)s(h')U_{h+h'}$ for $0.098 \leqslant U_hU_{h'} \leqslant 0.102$. The results are plotted graphically in Fig. 1 together with the theoretical curve for $U_hU_{h'} = 0.1$. The fit of the data is fairly good and agreement with the theoretical values of p will follow, since this is derived directly from the form of the function $P(U_{h+h'})$.

(ii) Comparison of theory and results for a real structure

Cochran (1952) published figures showing the efficiency of the sign relationships for glutamine. The

Table 1. *Test of theory*

$\bar{U}$	$\bar{\gamma}$	$2\sqrt{\varepsilon}.(\bar{\gamma})^3$	Q	P	Experimental		Theory	
					holds	fails	holds	fails
0·20–0·24	1·375	0·832	2·30	0·697	7	4	8	3
0·25–0·29	1·687	1·539	4·66	0·823	43	12	45	10
0·30–0·34	2·000	2·560	12·93	0·928	38	5	40	3
0·35–0·39	2·312	3·955	52·16	0·981	19	0	19	0
0·40–0·44	2·625	5·786	326·10	0·997	7	0	7	0

results are given in the form of the number of times the sign relationship holds and fails for various values of $\bar{U}$, the mean value of the three structure factors involved.
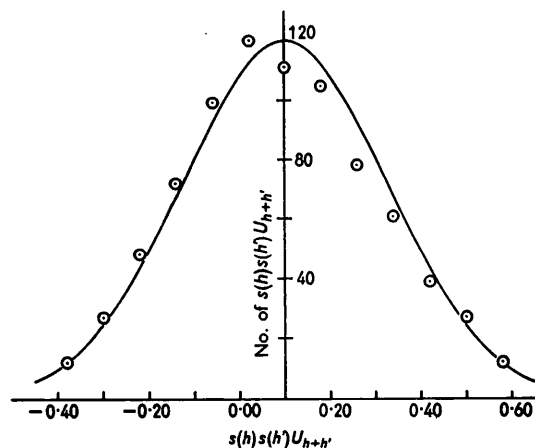

Fig. 1.

Glutamine, $C_5O_3N_2H_{10}$, contains four molecules per unit cell (space group $P2_12_12_1$). The scattering factors of carbon, nitrogen and oxygen are assumed to have the constant ratio $1\cdot0:1\cdot2:1\cdot4$, and the value of $\sqrt{\varepsilon}\ (= (\overline{U^2})^{\frac{1}{2}})$ is found to be 0·160 (the hydrogen atoms are not considered). The values of $\bar{U}$ given by Cochran are used to find $\bar{\gamma}$, the value of γ for the mean $\bar{U}$ of the range, which is used for calculating the theoretical figures (Table 1).

The above comparison is uncertain because of the following: (a) a few of the low-order U's were over-estimated by 20% (Cochran, 1952); (b) two atoms (carbon and oxygen) overlap in the $h0l$ projection; (c) $|\bar{U}^3| \geqslant |U_hU_{h'}U_{h+h'}|$, where $\bar{U}$ is the mean value of the three unitary structure factors. The agreement is nevertheless good enough to be a reliable indication of the number of successes to be expected.

References

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65.
CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton: University Press.
GILLIS, J. (1948). *Acta Cryst.* **1**, 76.
HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70.
KARLE, J. & HAUPTMANN, H. (1950). *Acta Cryst.* **3**, 181.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60.
WILSON, A. J. C. (1948). *Acta Cryst.* **2**, 318.
ZACHARIASEN, W. H. (1952). *Acta Cryst.* **5**, 68.