Acta Cryst. (1972). B28, 3362

Exponential Form of Joint Probability Distribution, Inequalities and Expected Values

BY J. KARLE

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20390, *U.S.A.*

(*Received* 4 *April* 1972)

The joint probability distributions for three structure factors whose subscripts add to zero is expressed in an exponential form for space groups $P1$ and $P\bar{1}$. These space groups serve as examples of the non-centrosymmetric and centrosymmetric cases, respectively. The exponential form effects considerably improved convergence properties although the behavior remains asymptotic. For the range of values of the normalized structure-factor magnitudes ordinarily obtained in experiment, the exponential forms are quite accurate. However, the accuracy is least for the largest possible values of the structure-factor magnitudes. By introducing a result from the inequality theory, which is most definitive when these magnitudes are largest, it is possible to alter the functional form of the exponential series to obtain joint probability distribution functions which are accurate over the entire range of the structure-factor magnitudes. Several probability measures of interest such as expected values and the probability that a structure factor has a positive sign are derived from the joint distribution functions.

Introduction

The joint probability distribution (Cramér, 1971, pp. 154 ff.) is a mathematical device for describing the dependent variation of several variates. This type of function is well suited for developing useful relationships among the crystal structure factors, the variates, which are mutually dependent upon the atomic coordinates, and was introduced into crystallographic theory for this purpose. It was recognized early that the simple inequality relationships which showed promise as aids in phase determination had, in fact, additional significance in the probability sense which extended beyond the range in which the inequalities held strictly. It was therefore apparent that the investigation of the probabilistic aspects of the inequality relationships would extend the range of their usefulness.

A consequence of the application of the joint probability distribution to structure factors was the derivation of certain phase determining formulas whose validity could be evaluated by probability measures. Some of these formulas corresponded to previous, explicitly stated inequalities and others were new. For example, the $\sum_1$ formula (Hauptman & Karle, 1953*a*) corresponds to certain inequalities of Harker & Kasper (1948, p. 72) and the $\sum_2$ formula (Hauptman & Karle, 1953*a*) corresponds to inequality (34) of Karle & Hauptman (1950). Formulas such as the $\sum_3$ type (Hauptman & Karle, 1953) also have correspondences in the inequality theory. The $B_{3,0}$ formula (Hauptman & Karle, 1958; Karle & Hauptman, 1958) appeared as a generalization of a formula already obtained from an analysis of the superposition of Patterson maps (Vaughan, 1958) and also by algebraic means (Hauptman & Karle, 1957; Karle & Hauptman, 1957).

The joint probability distributions also afforded a measure of the validity of the phase-determining formulas, but because they were expressed in series form the measures proposed were not as convenient or accurate to use as those obtained from the central limit theorem. By comparing the series from the joint distribution for centrosymmetric crystals (Hauptman & Karle, 1953*a*) and for non-centrosymmetric ones (Karle & Hauptman, 1956) with the results from the central limit theorem (Woolfson, 1954; Cochran & Woolfson, 1955; Cochran, 1955), it is apparent that the expressions become comparable if a manipulation which places an appropriate term, x, in the series into exponential form is employed, *i.e.* if the relation $1+x \sim \exp(x)$ is used. This point is of more than passing interest, since it will be seen that the transformation of the entire series expansions for the joint probability distributions to an exponential form can improve significantly their accuracy and interpretation. The advantage of the exponential form has been described by Bertaut (1960) for centrosymmetric crystals, who noted that in this form series terms of order $N^{-m/2}$, where N is the number of atoms in the unit cell. are associated with structure-factor polynomials whose highest degree is $m+2$. This also holds for non-centrosymmetric ones. It will be seen that this improves the convergence (aside from numerical coefficients) by a factor of the order of $|E|^{2m-2}$, where $|E|$ represents a normalized structure factor magnitude.

As an additional illustration of the probability aspects of the inequalities, it has recently been shown (Karle, 1971) that probability measures with improved expressions for the variance applicable to the $\sum_2$ and tangent formulas can be read directly from a form of the inequality theory given by expression (30) of Karle & Hauptman (1950) using the central limit theorem. Similar probability measures and additional ones to be associated with the inequalities have also been derived by Tsoucaris (1970).

After the initial applications of the joint probability distribution to the problem of phase determination (Hauptman & Karle, 1953*a*; Karle & Hauptman, 1956) several authors (Klug, 1958; Naya, Nitta & Oda, 1964, 1965) discussed the higher-order terms in the series expansion, giving particular emphasis to the possibility of increased accuracy. This possibility may be rather elusive for the series expansion when statistical information involving the larger normalized structure-factor magnitudes is sought. The situation is not necessarily improved by increasing the number of higher-order terms considered. The reason for this is that the series are asymptotic, *i.e.* they do not ultimately converge as the number of terms increases without limit, but rather approach the true value of the function being represented up to some term, after which they diverge. In this context, it is often both prudent and necessary in problems of interest to crystallography involving the largest normalized structure-factor magnitudes to accept no more terms in the series than the first one in which the quantity of interest appears.

Because of the varying mathematical terminologies and the alternative book-keeping of the moment calculations employed by the various authors, it might appear on a first reading of the papers on the joint probability distribution that the formulation of the problem and the results differ from one to the other. It is worth noting that in each case the formulation of the joint probability is the same and the statistical model involving the uniform distribution of atomic coordinates in the unit cell is the same. An exception is the derivation of the $B_{3,0}$ formula in which the independent random variable is assumed to be uniformly distributed over reciprocal space. A further development of this latter approach has been recently presented by Hauptman (1971). The identity of the formulation is reflected in the identity of the results, except for the number of terms recorded. For example, the results for centrosymmetric crystals are characteristically expressed in terms of the Hermite polynomials. This may be seen in the results of Hauptman & Karle (1953*a*) of Bertaut (1955), of Klug (1958) and those of Naya, Nitta & Oda (1964). They also appear in the probability distribution for a single centrosymmetric structure factor (Karle & Hauptman, 1953*a*). For non-centrosymmetric crystals, the results are characteristically expressed in terms of associated Laguerre polynomials, as can be seen for the joint distribution appropriate to the real and imaginary parts of a single structure factor (Hauptman & Karle, 1953*b*). They are also contained in the formalism of Bertaut (1956) of Karle & Hauptman (1956), and of Naya, Nitta & Oda (1965).

As mentioned, the joint probability distribution can be expressed as an exponential function in which the argument of the function is a series in negative powers of the number of atoms in the unit cell, similar to the usual series expression. The series in the exponential form are still asymptotic in character, but their asymptotic convergence properties are considerably improved. This affords the opportunity to obtain improved estimates for statistical quantities such as expected values and variances. Of particular interest are the invariants $\varphi_{\mathbf{h}}-\varphi_{\mathbf{k}}-\varphi_{\mathbf{h}-\mathbf{k}}$ and the cosines of these invariants because of their importance in phase determining procedures. In attempting to derive a general expression for the expected value of a cosine invariant, $\cos(\varphi_{\mathbf{h}}-\varphi_{\mathbf{k}}-\varphi_{\mathbf{h}-\mathbf{k}})$, from the joint distribution, it cannot be expected that an asymptotic series would afford a function which would be valid for the largest values of the normalized structure factor magnitudes. Nevertheless, it is probable that a strong clue regarding the form of this function can be obtained from the joint distribution. Coupling this information with information obtainable from the low-order inequalities, namely, relations which are strictly valid only for the very largest values of the normalized structure-factor magnitudes, can conceivably provide a reliable indication regarding the appropriate functional form for the desired statistical quantity. Thus information from the joint distribution, whose asymptotic behavior generates a deterioration of accuracy when applied to the largest normalized structure-factor magnitudes, would be joined to information from the inequality theory which behaves best when applied to these large structure factor magnitudes.

The series expansions of the joint distribution for $E_{\mathbf{h}}, E_{-\mathbf{k}}, E_{-\mathbf{h}+\mathbf{k}}$ for centrosymmetric crystals and for $|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}-\mathbf{k}}|$, $\varphi_{\mathbf{h}}, \varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}}$ for noncentrosymmetric ones have been derived to the $N^{-5/2}$ term (Naya, Nitta & Oda, 1964, 1965), where N is the number of atoms in the unit cell. The exponential forms of these distributions for the equal-atom case, their properties and relation to statistical quantities of interest will be described with a view toward obtaining more precise information regarding the cosine invariants. Effectively, this means making more accurate use of the known values for the normalized structure-factor magnitudes associated with the $\varphi_{\mathbf{h}}$, $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h}-\mathbf{k}}$. In terms of reducing the degree of the polynomials in $|E|$ associated with the $N^{-m/2}$ term by $|E|^{2m-2}$, it is seen when $m=5$, polynomials in $|E|$ are reduced by $|E|^8$, a considerable reduction when the $|E|$ are large.

Exponential form for joint distribution noncentrosymmetric reflections ($\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=\mathbf{0}$)

The joint probability distribution function for three complex-valued normalized structure factors, $E_1=E_{\mathbf{h}_1}$, $E_2=E_{\mathbf{h}_2}$, $E_3=E_{\mathbf{h}_3}$, on the assumption that the atomic coordinates are random variables which are uniformly and independently distributed, is considered now for the case of equal atoms in space group $P1$. Formalisms for obtaining joint distribution functions for noncentrosymmetric reflections have been described (Hauptman & Karle, 1953*b*; Bertaut, 1956; Karle & Hauptman, 1956; Naya, Nitta & Oda, 1965). The joint distribution function for E_1, E_2, E_3 may be written in the exponential form

$$P(|E_1|,|E_2|,|E_3|,\varphi_1,\varphi_2,\varphi_3)=\frac{1}{\pi^3}|E_1E_2E_3|$$
$$\times\exp(-|E_1|^2-|E_2|^2-|E_3|^2)$$
$$\times\exp\left[\frac{2}{N^{1/2}}|E_1E_2E_3|\left\{1+\frac{1}{N}(|E_1|^2+|E_2|^2\right.\right.$$
$$+|E_3|^2-3)+\frac{1}{N^2}([|E_1|^4+(11/4)|E_1E_2|^2$$
$$\left.\left.-(9/2)|E_1|^2+\text{cyc.}]+2)\right\}\cos(\varphi_1+\varphi_2+\varphi_3)\right.$$
$$-\frac{5}{4N^2}|E_1E_2E_3|^2\cos 2(\varphi_1+\varphi_2+\varphi_3)$$
$$-\frac{1}{N}\{[(1/4)|E_1|^4+|E_1E_2|^2-2|E_1|^2+\text{cyc.}]+5/2\}$$
$$-\frac{1}{N^2}\{[(5/36)|E_1|^6+|E_1|^4|E_2|^2+|E_1|^2|E_2|^4+\text{cyc.}]$$
$$+4|E_1E_2E_3|^2-[(9/8)|E_1|^4$$
$$\left.+4|E_1E_2|^2-(3/2)|E_1|^2+\text{cyc.}]+5/8\}+\ldots\right]. \quad (1)$$

By expanding the second exponential function on the the right side, equation (1) is seen to be identical with equation (67) of Naya, Nitta & Oda (1965) for the case of equal atoms (the coefficient of the entire Z_4Z_5 term should be $\frac{1}{4}$, *i.e.* the 2 should be replaced by $\frac{1}{4}$, and that of the $Z_5^2Z_5$ term should be $\frac{3}{2}$, *i.e.* the 1 should be replaced by $\frac{3}{2}$). The initial factor $|E_1E_2E_3|$ of equation (1) is understood to be inserted into the latter equation.

An expected value of considerable practical importance is the expected value of the cosine invariant, $\cos(\varphi_1+\varphi_2+\varphi_3)$, given the known values of $|E_1|$, $|E_2|$, $|E_3|$. This circumstance is the one commonly met in practice in which the magnitudes of the structure factors are known, but the phases are not. In order to determine this expected value, the conditional distribution of the sum $\varphi_1+\varphi_2+\varphi_3$ given $|E_1|$, $|E_2|$ and $|E_3|$ must be obtained.

If we define

$$\Phi_{123}=\varphi_1+\varphi_2+\varphi_3\,, \quad (2)$$

the desired conditional distribution $P_1(\Phi_{123};|E_1|,|E_2|,|E_3|)$ is obtained from (1) by fixing the values of $|E_1|$, $|E_2|$ and $|E_3|$ and renormalizing. This gives

$$P_1(\Phi_{123};|E_1|,|E_2|,|E_3|\simeq K_1$$
$$\times\exp\left[\frac{2}{N^{1/2}}|E_1E_2E_3|p(|E_1|,|E_2|,|E_3|)\right.$$
$$\left.\times\cos\Phi_{123}-\frac{5}{4N^2}|E_1E_2E_3|^2\cos 2\Phi_{123}\right], \quad (3)$$

where K_1, the normalizing constant, is found to be

$$K_1=\left[2\pi I_0\left(\frac{2|E_1E_2E_3|p}{N^{1/2}}\right)\right.$$
$$\left.-\frac{5\pi|E_1E_2E_3|^2}{2N^2}I_2\left(\frac{2|E_1E_2E_3|p}{N^{1/2}}\right)\right]^{-1}, \quad (4)$$

the I_n are Bessel functions of imaginary argument and

$$p=p(|E_1|,|E_2|,|E_3|)=1+\frac{1}{N}(|E_1|^2+|E_2|^2+|E_3|^2-3)$$
$$+\frac{1}{N^2}([|E_1|^4+(11/4)|E_1E_2|^2-(9/2|E_1|^2+\text{cyc.}]+2) \quad (5)$$

or

$$p=1-(3/N)+(2/N^2)+|U_1|^2+|U_2|^2+|U_3|^2$$
$$+[|U_1|^4+(11/4)|U_1U_2|^2-(9/2N)|U_1|^2+\text{cyc.}]\,. \quad (6)$$

The expected value for $\cos\Phi_{123}$ given $|E_1|$, $|E_2|$, $|E_3|$, is obtained from

$$\langle\cos\Phi_{123}\rangle\simeq\int_0^{2\pi}\cos\Phi_{123}P_1(\Phi_{123};|E_1|,|E_2|,|E_3|)\mathrm{d}\Phi_{123}\,. \quad (7)$$

Evaluation of equation (7) gives

$$\langle\cos\Phi_{123}\rangle\simeq$$
$$\frac{\left(1-\frac{5|E_1E_2E_3|^2}{8N^2}\right)\frac{I_1(w)}{I_0(w)}-\frac{5|E_1E_2E_3|^2}{8N^2}\frac{I_3(w)}{I_0(w)}}{1-\frac{5|E_1E_2E_3|^2}{4N^2}\frac{I_2(w)}{I_0(w)}} \quad (8)$$

where

$$w=\frac{2|E_1E_2E_3|p}{N^{1/2}}\,. \quad (9)$$

The expected value given in equation (8) is defined as far as the $N^{-5/2}$ term, similar to the probability distribution (3) from which it was derived.

The probability distribution functions (1) and (3) and the expected value for $\cos(\varphi_1+\varphi_2+\varphi_3)$ in (8) should afford good accuracy for the magnitudes of $|E|$ normally encountered in crystals. However, it is apparent from examining these expressions that since the maximum value for $|E|$ is $N^{1/2}$, though very rarely closely attained, the accuracy can deteriorate for very large $|E|$ values. It is understood that accuracy in calculating asymptotic series is optimized by employing the series up to and including the last converging term, when possible, and ignoring the remaining ones. In the polynomial p, equation (5), the degree of the products of $|E|$ associated with $N^{m/2}$ is m, rather than $m+2$ which occurs in other terms in the probability expressions. The convergence properties of p should be satisfactory for most structures of interest.

By altering the probability function (3) according to a suggestive indication arising from the inequality theory, it is possible to obtain a probability function which is quite accurate for all possible values of the

$|E|$. The inequality of interest is the one of third order (Karle & Hauptman, 1950),

$$\begin{vmatrix} 1 & U_{-\mathbf{k}} & U_{-\mathbf{h}} \\ U_{\mathbf{k}} & 1 & U_{-\mathbf{h}+\mathbf{k}} \\ U_{\mathbf{h}} & U_{\mathbf{h}-\mathbf{k}} & 1 \end{vmatrix} \geq 0 \qquad (10)$$

which, when expanded, can be written

$$\cos(\varphi_{\mathbf{h}}-\varphi_{\mathbf{k}}-\varphi_{\mathbf{h}-\mathbf{k}}) \geq \frac{|U_{\mathbf{h}}|^2+|U_{\mathbf{k}}|^2+|U_{\mathbf{h}-\mathbf{k}}|^2-1}{2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|} . \qquad (11)$$

This inequality (11) affords an insight into how the large values of the $|U|$ impose a constraint on the cosine invariant. It is apparent, for example, that the cosine invariant is positive if $|U_{\mathbf{h}}|^2+|U_{\mathbf{k}}|^2+|U_{\mathbf{h}-\mathbf{k}}|^2>1$ where each $|U|>0$. The maximum value of the right side of (11) is unity. This is, of course, attained when each $|U|=1$, its maximum value. However, in the context of the inequality, the maximum value is more easily attained. For example if one of the three $|U|=1$, the other two must be equal and the right side then is equal to unity. Unequal values for the other two $|U|$ would violate the inequality.

The polynomial in equation (5) is now approximated for modest values of $|U|=E/N^{1/2}$, *i.e.* $|U|<0{\cdot}2$, and moderately large N, *i.e.* $N>100$, with

$$p \simeq \frac{1}{1-|U_1|^2-|U_2|^2-|U_3|^2} . \qquad (12)$$

If this expression for p were substituted into equation (3) and higher order terms were neglected, the exponential in (3) would become

$$\exp\left[\frac{2|E_1E_2E_3|}{N^{1/2}(1-|U_1|^2-|U_2|^2-|U_3|^2)}\cos\Phi_{123}\right] . \qquad (13)$$

Although the denominator of the argument of (13) is satisfactory for small values for the $|U|$, it is evidently quite unsatisfactory for the larger values since it approaches a minimum value of $-2N^{1/2}$ instead of the value of zero, the latter implying certainty or zero variance to be associated with the largest values for the $|U|$.

Recalling that the right side of (11) approaches unity as the $|U|$ approach their maximum value, we consider taking one minus this function to find a more appropriate function for the denominator of the argument of (13). This becomes

$$\left[1-\frac{|U_{\mathbf{h}}|^2+|U_{\mathbf{k}}|^2+|U_{\mathbf{h}-\mathbf{k}}|^2-1}{2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|}\right]^{-1} = \frac{2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|}{1+2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|-|U_{\mathbf{h}}|^2-|U_{\mathbf{k}}|^2-|U_{\mathbf{h}-\mathbf{k}}|^2} . \qquad (14)$$

The observation that the denominator of (14) approaches zero for the largest values of $|U|$, while the term $2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|$ has little effect for small values of $|U|$ and a comparison with the argument of (13) suggest that a new, accurate probability distribution function may be written,

$$P_2(\Phi_{123};|E_1|,|E_2|,|E_3|) \simeq K_2 \times \exp\left[\frac{2|E_1E_2E_3|}{N^{1/2}q_i}\cos\Phi_{123}\right] , \quad i=1 \text{ or } 2 \qquad (15)$$

where

$$K_2 = \left[2\pi I_0\left(\frac{2|E_1E_2E_3|}{N^{1/2}q_i}\right)\right]^{-1} \qquad (16)$$

and

$$q_1 = 1+2|U_1U_2U_3|-|U_1|^2-|U_2|^2-|U_3|^2. \qquad (17)$$

An alternative form for q_1 which is a good approximation for crystallographic data is

$$q_2 = (1-|U_1|^2)(1-|U_2|^2)(1-|U_3|^2) . \qquad (18)$$

The desired expression for the expected value of the cosine invariant from P_2 can be obtained from a relation similar to equation (7), giving

$$\langle\cos(\varphi_1+\varphi_2+\varphi_3)\rangle \simeq \frac{I_1\left(\frac{2|E_1E_2E_3|}{N^{1/2}q_i}\right)}{I_0\left(\frac{2|E_1E_2E_3|}{N^{1/2}q_i}\right)} , \quad i=1 \text{ or } 2 . \qquad (19)$$

The preferable form for q can be determined by numerical tests.

The probablity distribution (1) modified in a comparable fashion to (15) can also express the modified probability distribution, P_3, for a single phase, φ_1, given φ_2, φ_3, $|E_1|$, $|E_2|$, $|E_3|$ or the modified probability distribution, P_4, for the sum of two phases $\varphi_2+\varphi_3$, given φ_1, $|E_1|$, $|E_2|$, $|E_3|$. In fact, since the range of variation is $-\pi<\varphi\leq\pi$, where φ represents φ_1 or $\varphi_2+\varphi_3$, P_3 and P_4 are both equal to P_2, each having the same value for the normalizing constant K_2. We may write

$$P_2(\Phi_{123};|E_1|,|E_2|,|E_3|) = P_3(\varphi_1;\varphi_2,\varphi_3,|E_1|,|E_2|,|E_3|) = P_4(\varphi_2+\varphi_3;\varphi_1,|E_1|,|E_2|,|E_3|) . \qquad (20)$$

These probability formulas can be compared to that of Cochran [1955, equation (7)] which is obtained when the $q_i=1$ or, effectively, when the $|U_i|$ are small.

The alteration of P_3 to obtain the probability distribution for φ_1, given many sets of φ_2, φ_3, $|E_2|$, $|E_3|$ satisfying $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$, can be obtained by multiplying the individual probability distributions of type P_3 together, assuming that the distributions are independent, and renormalizing. This type of manipulation to give a probability distribution and variance measure based on Cochran's formula appears in equations (3·25) and (3·33)* respectively, of Karle & Karle (1966). These two formulas, along with additional mathematical expressions in sections 3·2, 3·3 and 3·4 of the latter reference, can be quite readily altered to be consistent with

* Note that the plus sign before the last term of this formula should be minus. However, Fig. 2 was calculated correctly.

the results of this paper by simply redefining the symbol κ of equation (3·20) by dividing by q_i to give

$$\kappa(\mathbf{h},\mathbf{k})=2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|/q_i \qquad (21)$$

where we identify $\mathbf{h}_1=\mathbf{h}$, $\mathbf{h}_2=-\mathbf{k}$ and $\mathbf{h}_3=-\mathbf{h}+\mathbf{k}$. For equal atoms, $\sigma_3\sigma_2^{-3/2}=N^{1/2}$.

A probability distribution for $\varphi_{\mathbf{h}}$ has recently been derived using the central limit theorem (Karle, 1971) which can be compared to (15) and (18). It is the same except for the first factor in (18) which is replaced by unity. From the point of view of the central limit theorem, expressions (17) and (18) are proportional to the variance of the real and imaginary parts of a structure factor $E_{\mathbf{h}_1}$, given $E_{\mathbf{h}_2}$ and $E_{\mathbf{h}_3}$ where $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$. This matter is discussed further in the Appendix where it is shown that the expression for the variance used previously (Karle, 1971), based on the square of the radius of a bounding circle in the complex plane for the unitary structure factor $U_{\mathbf{h}}$, is somewhat more conservative than equation (17) and its generalizations.

Exponential form for joint distribution centrosymmetric reflections ($\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$)

The joint probability distribution function for three normalized structure factors, $E_1=E_{\mathbf{h}_1}$, $E_2=E_{\mathbf{h}_2}$, $E_3=E_{\mathbf{h}_3}$, on the assumption that the atomic coordinates are random variables which are uniformly and independently distributed, is considered for the case of equal atoms in space group $P\bar{1}$. Formalisms for obtaining joint distribution functions for centrosymmetric reflections have been described (Hauptman & Karle, 1953; Bertaut, 1955; Klug, 1958; Naya, Nitta & Oda, 1964). Bertaut (1960) has discussed several examples of the exponential form for the joint distribution for centrosymmetric reflections. The exponential form of the joint distribution for E_1, E_2, E_3 valid to the $N^{-5/2}$ term may be written:

$$\begin{aligned}P(E_1,E_2,E_3)=&\frac{1}{(2\pi)^{3/2}}\exp\left[-\tfrac{1}{2}(E_1^2+E_2^2+E_3^2)\right]\\&\times\exp\left[\frac{1}{N^{1/2}}\left\{1+\frac{1}{N}(E_1^2+E_2^2+E_3^2-4)\right.\right.\\&+\frac{1}{N^2}[E_1^4+(11/4)E_1^2E_2^2-6E_1^2+\text{cyc.}]\\&\left.\left.+11/4\right\}E_1E_2E_3-\frac{1}{N}\{[(1/8)E_1^4+(1/2)E_1^2E_2^2\right.\\&-(5/4)E_1^2+\text{cyc.}]+13/8\}-\frac{1}{N^2}\{[(5/72)E_1^6\\&+(1/2)E_1^4E_2^2+(1/2)E_1^2E_2^4+(21/8)E_1^2E_2^2E_3^2\\&-(35/48)E_1^4-(21/8)E_1^2E_2^2+(9/8)E_1^2+\text{cyc.}]\\&\left.+5/8\}\ldots\right]. \qquad (22)\end{aligned}$$

By expanding the second exponential function on the right side, equation (22) is seen to be identical with equation (III-2) of Naya, Nitta & Oda (1964).

Of interest are questions which arise in practice, *i.e.* the probability that the sign of $E_1E_2E_3$ is positive, given $|E_1|$, $|E_2|$ and $|E_3|$, or the probability that the sign of E_1 is positive, given $|E_1|$, E_2 and E_3. Since the analyses leading to the desired results are quite similar, the steps will be outlined for the first case and only the answer will be written for the second. The following algebra is employed:

$$P_+ + P_- = 1 \qquad (23)$$

$$P_+ = (P_+/P_-)/[1+(P_+/P_-)]\,, \qquad (24)$$

where P_+ is the probability that some quantity is positive. The ratio of the probability that a quantity x is positive, $P_+(x)$, to the probability that it is negative, $P_-(x)$, is given by the ratio of the values of the probability distribution function for the positive and negative values of x. If the probability distribution function, P, is proportional to an exponential function of x, *e.g.* $P\propto e^{ax}$, then $P_+(x)/P_-(x)=\exp(a|x|)/\exp(-a|x|)=\exp(2a|x|)$ and, employing (24) it follows that

$$P_+(x)=\exp(2a|x|)/[1+\exp(2a|x|)] \qquad (25)$$

and

$$P_+(x)=\tfrac{1}{2}+\tfrac{1}{2}\tanh(a|x|)\,. \qquad (26)$$

The probability that the sign of $E_1E_2E_3$ is plus, given $|E_1|$, $|E_2|$ and $|E_3|$, $P_+(E_1E_2E_3;|E_1|,|E_2|,|E_3|)$, becomes from (22) and (26),

$$P_+(E_1E_2E_3;|E_1|,|E_2|,|E_3|)=\tfrac{1}{2}+\tfrac{1}{2}\tanh\frac{1}{N^{1/2}}p_1|E_1E_2E_3| \qquad (27)$$

where

$$\begin{aligned}p_1=&1+\frac{1}{N}(E_1^2+E_2^2+E_3^2-4)\\&+\frac{1}{N^2}\{[E_1^4+(11/4)E_1^2E_2^2-6E_1^2+\text{cyc.}]+11/4\} \qquad (28)\end{aligned}$$

or

$$\begin{aligned}p_1=&1-4/N+11/4N^2+U_1^2+U_2^2+U_3^2\\&+[U_1^4+(11/4)U_1^2U_2^2-(6/N)U_1^2+\text{cyc.}]\,. \qquad (29)\end{aligned}$$

The probability that the sign of E_1 is positive, given $|E_1|$, E_2 and E_3, $P_+(E_1;|E_1|,E_2,E_3)$, is

$$P_+(E_1;|E_1|,E_2,E_3)=\tfrac{1}{2}+\tfrac{1}{2}\tanh\frac{1}{N^{1/2}}p_1|E_1|E_2E_3\,. \qquad (30)$$

The polynomial p_1, equation (28), should have good convergence properties for most crystals of interest.

The question again arises as to whether a suitable alteration can be made on the polynomial p_1, occurring as part of the coefficient of $E_1E_2E_3$ in (22), in order to obtain accurate measures of P_+ throughout the possible range of values for the $|E|$. Drawing again on the

third-order inequality (10), specialized to the real structure factors of centrosymmetric crystals, (11) is replaced by

$$s_{\mathbf{h},\mathbf{k}} \geq \frac{U_{\mathbf{h}}^2+U_{\mathbf{k}}^2+U_{\mathbf{h}-\mathbf{k}}^2-1}{2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}|}, \tag{31}$$

where $s_{\mathbf{h},\mathbf{k}}$ is a two-valued function having the values ± 1, and representing the sign of $U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}$. This inequality has been previously discussed by Klug (1958) who noted that the sign would have to be positive if $U_{\mathbf{h}}^2+U_{\mathbf{k}}^2+U_{\mathbf{h}-\mathbf{k}}^2>1$. There is a further restriction which is somewhat less apparent, namely, there are combinations of values for $|U_{\mathbf{h}}|$, $|U_{\mathbf{k}}|$ and $|U_{\mathbf{h}-\mathbf{k}}|$ when $U_{\mathbf{h}}^2+U_{\mathbf{k}}^2+U_{\mathbf{h}-\mathbf{k}}^2>0{\cdot}75$ for which $s_{\mathbf{h},\mathbf{k}}$ must be $+1$. This occurs because for certain combinations of values for $|U_{\mathbf{h}}|$, $|U_{\mathbf{k}}|$ and $|U_{\mathbf{h}-\mathbf{k}}|$, the magnitude of the denominator of (31) exceeds that of the numerator. Under such circumstances, $s_{\mathbf{h},\mathbf{k}}=-1$ would violate the inequality. An example is $|U_{\mathbf{h}}|=|U_{\mathbf{h}-\mathbf{k}}|=0{\cdot}51$.

By using arguments which are similar to those applied to the noncentrosymmetric case, it is suggested that good accuracy may be obtained over the full range of possible values for the $|E|$ if $q_i (i=1$ or $2)$, defined in equations (17) and (18), were to replace p_1 in equations (27) and (30). For example, equation (30) would be replaced by

$$P_+(E_1;|E_1|,E_2,E_3)=\tfrac{1}{2}+\tfrac{1}{2}\tanh\frac{|E_1|E_2E_3}{N^{1/2}q_i},\ i=1 \text{ or } 2\,. \tag{32}$$

When there are several sets of known $E_{\mathbf{h}_2}$ and $E_{\mathbf{h}_3}$ such that $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$, equation (30) becomes

$$P_+(E_{\mathbf{h}})=\tfrac{1}{2}+\tfrac{1}{2}\tanh\frac{|E_{\mathbf{h}}|\sum_{\mathbf{k}}(E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}/q_i)}{N^{1/2}},\quad i=1 \text{ or } 2\,, \tag{33}$$

where $\mathbf{h}_1=\mathbf{h}$, $\mathbf{h}_2=\mathbf{k}$ and $\mathbf{h}_3=\mathbf{h}-\mathbf{k}$. It is seen that the hyperbolic tangent formula (33) is the same as the one commonly used (Woolfson, 1954; Cochran & Woolfson, 1955) if $q_i=1$.

Numerical tests

Some numerical tests have been carried out in order to compare several formulas for the expected value of the cosine invariants and to obtain an estimate of their average error. Results are shown in Table 1 for a random structure in space group $P1$ composed of 50 equal atoms and in Table 2 for solaphyllidine.CH_3OH (Karle, 1970) which crystallizes in space group $P2_12_12_1$ with 148 almost equal nonhydrogen atoms and 204 hydrogen atoms in the unit cell. Examination of the tables shows the distinction between the calculated averages of expected values obtained when q_i is equal to q_1 or q_2 or when it is replaced by unity. The latter corresponds to using only the first term in the joint distribution function. The difference between the use of q_1 or q_2 is small. They both lead to higher results for the expected values than when q_i is replaced by unity. This, of course, becomes more significant as the $|E_1E_2E_3|$ increase in value.

Additional calculations of the expected values were performed which were based on the higher order terms

Table 1. *Calculations of the average of a number of cosine invariants for* 50*-atom random structure in space group* $P1$ *using equation* (19)

Sample	$[(\cos\Phi_{123})_{obs}]_{av}$	$[\langle\cos\Phi_{123}\rangle_{calc}]_{av}$	q_i in equation (19)	Number of contributions	Average error
All	0·730	0·678	1		0·28
$\|E\|>1{\cdot}7$	0·730	0·745	q_1	540	0·25
	0·730	0·754	q_2		0·25
All	0·789	0·750	1		0·23
$\|E\|>1{\cdot}85$	0·789	0·811	q_1	96	0·21
	0·789	0·819	q_2		0·20
All	0·772	0·805	1		0·22
$\|E\|>2{\cdot}0$	0·772	0·856	q_1	21	0·20
	0·772	0·866	q_2		0·19

Table 2. *Calculation of the average of a number of cosine invariants for solaphyllidine, having* 148 *almost equal nonhydrogen atoms in the unit cell of space group* $P2_12_12_1$, *using equation* (19)

Sample	$[(\cos\Phi_{123})_{obs}]_{av}$	$[\langle\cos\Phi_{123}\rangle_{calc}]_{av}$	q_i in equation (19)	Number of contributions	Average error
All	0·570	0·540	1		0·43
$\|E\|>1{\cdot}7$	0·570	0·564	q_1	254	0·42
	0·570	0·565	q_2		0·42
All	0·626	0·599	1		0·37
$\|E\|>1{\cdot}85$	0·626	0·624	q_1	109	0·36
	0·626	0·625	q_2		0·36
All	0·663	0·663	1		0·34
$\|E\|>2{\cdot}0$	0·663	0·688	q_1	44	0·33
	0·663	0·690	q_2		0·33

in equation (1), *e.g.* as given by equation (8). For the present examples no significant differences were obtained from the results given by equation (19).

The third test in Table 1 shows poorer agreement between the observed and calculated averages of the cosine invariants than the first two. It is seen that only 21 invariants with all $|E|>2{\cdot}0$ were used in the comparison, a rather small sample. Nevertheless, as shown in the last column of Table 1, the average deviation between the observed values for the individual invariants and the expected values calculated from equation (19) is small.

Regularities in a structure could affect, particularly, the low-order reflections and could cause a discrepancy between the average of the observed values for the cosine invariants and the theoretical estimates.

Concluding remarks

A main feature of the exponential form for the joint probability distribution function is its improved convergence properties and the opportunity this affords to obtain accurate measures of a variety of statistical properties. An effect of this is to enhance the expected value of a cosine invariant and the probability that the sign of a centrosymmetric reflection is plus compared to earlier measures. With the aid of inequality theory, these statistical measures are expressed in terms of the q_i which replace the polynomials p and p_1, equations (19), (32) and (33). The q_i attain a maximum value of unity only when $|U_1|=|U_2|=|U_3|=0$ and decrease toward zero as the values of the $|U|$ increase.

In the mathematical analysis presented, only equal atoms were considered. A manuscript is in preparation in which the mathematical analysis is applied to the case of unequal atoms.

I wish to express my thanks to Mr Stephen Brenner who programmed and carried out the numerical tests. I also wish to thank Dr R. Gilardi for verifying the coefficients in the joint probability distributions, equations (1) and (22).

APPENDIX

Geometric interpretation of formula (17)

The diagram in Fig. 1 represents a bounding circle of radius r centered at δ in the complex plane for the unitary structure factor $U_{\mathbf{h}}$ of known magnitude. This is the geometric representation of inequality (30) of Karle & Hauptman (1950),

$$|U_{\mathbf{h}}-\delta|\leq r \qquad (A1)$$

in which the elements of the determinants forming δ and r are here defined as unitary structure factors. Examination of Fig. 1 shows that the range of possible values for the phase of $U_{\mathbf{h}}$ is 2α. The equation of the bounding circle is

$$(x-x_\delta)^2+(y-y_\delta)^2=r^2 \qquad (A2)$$

where the real and imaginary parts of δ are x_δ and y_δ respectively. The equation of the circle formed by the magnitude $|U_{\mathbf{h}}|$ which intersects the bounding circle at points E and G is

$$x^2+y^2=|U_{\mathbf{h}}|^2. \qquad (A3)$$

From (*A*2) and (*A*3), the equation for the line containing the points of intersection E and G is

$$2x_\delta x+2y_\delta y=|U_{\mathbf{h}}|^2+|\delta|^2-r^2. \qquad (A4)$$

The equation of the line perpendicular to the segment EG and passing through the origin at O and the points F and δ is

$$x=x_\delta y/y_\delta\,. \qquad (A5)$$

Equations (*A*4) and (*A*5) intersect at point F. Solving for the coordinates of F can lead readily to the length of the segment OF and, using this, $\cos\alpha$ is found to be

$$\cos\alpha=\frac{|U_{\mathbf{h}}|+|\delta|^2-r^2}{2|U_{\mathbf{h}}\delta|}\,. \qquad (A6)$$

Expecting that the variance of the real and imaginary parts of $E_{\mathbf{h}}$ should be expressible in terms of the angle α, use is made of the relation $1-\cos\alpha=2\sin^2\alpha/2$ and equation (*A*6) to make the reasonable proposal that

$$2\times\text{Variance}=4|U_{\mathbf{h}}\delta|\sin^2\alpha/2=r^2-(|U_{\mathbf{h}}|-|\delta|)^2\,. \qquad (A7)$$

Note that Fig. 1 implies that $0\leq\alpha/2\leq\pi/2$.

It is of interest to examine the properties of the variance formula (*A*7). An alternative interpretation of $4|U_{\mathbf{h}}\delta|\sin^2\alpha/2$ can be made by referring to Fig. 1 and noting that it is equal to the products of the lengths of the two chords AB and CD. Further examination of Fig. 1 shows that when $|U_{\mathbf{h}}|$ is of such a length as to

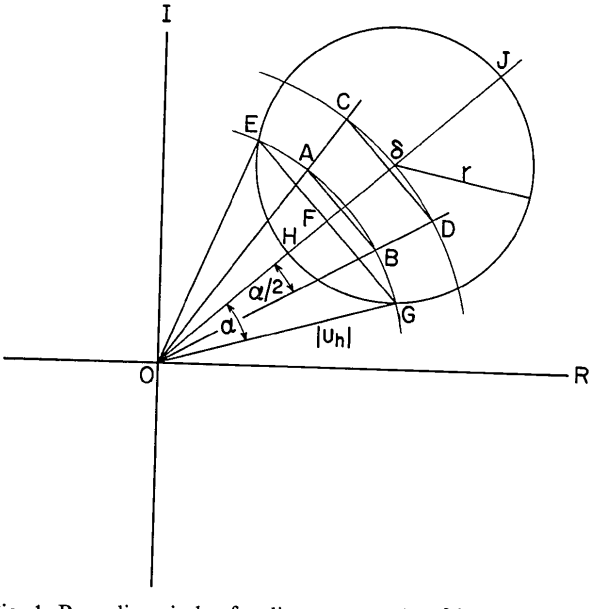

Fig. 1. Bounding circle of radius r centered at δ in the complex plane for the unitary structure factor $U_{\mathbf{h}}$.

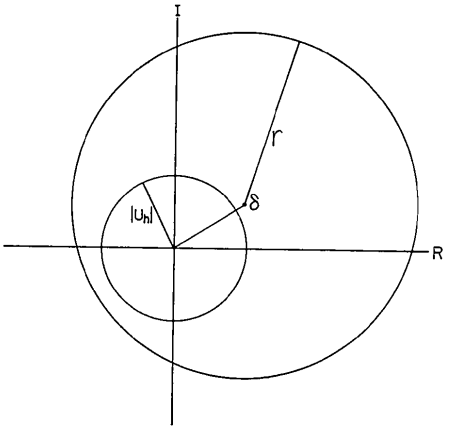

Fig. 2. Circle of radius $|U_{\mathbf{h}}|$ centered at the origin is contained within the bounding circle of radius r centered at δ in the complex plane when $|U_{\mathbf{h}}| < r - |\delta|$.

just touch the bounding circle either at points H or J, $\alpha = 0$ and there is no uncertainty concerning the value of the phase $\varphi_{\mathbf{h}}$. The variance given by (*A*7) is equal to zero under these circumstances.

When $|U_{\mathbf{h}}| = \delta$, the variance is $r^2/2$, the quantity suggested in a previous paper (Karle, 1971). Equation (*A*7) suggests that the variance varies between zero and $r^2/2$, depending upon the value of $|U_{\mathbf{h}}|$.

If the bounding circle generated by (*A*1) is based on a third-order determinant, $\delta = U_{\mathbf{k}} U_{\mathbf{h}-\mathbf{k}}$ and $r = (1 - |U_{\mathbf{k}}|^2)^{1/2}(1 - |U_{\mathbf{h}-\mathbf{k}}|^2)^{1/2}$. Substitution of these functions into (*A*7) gives equation (17).

There is a case for which the diagram in Fig. 1 does not apply. This occurs when the circle of radius $|U_{\mathbf{h}}|$ is contained within the bounding circle (Fig. 2), satisfying

$$|U_{\mathbf{h}}| < r - |\delta| \,. \qquad (A8)$$

Under these circumstances, the right side of (*A*7) is still retained to represent the variance, namely,

$$2 \times \text{Variance} = r^2 - (|U_{\mathbf{h}}| - |\delta|)^2. \qquad (A9)$$

When equations (17) and (18) are applied to the centrosymmetric case, *e.g.* when the q_i replace p_i in equations (28) and (30), they represent the variance and not twice the variance of a structure factor. The same would apply to (*A*9), the generalization of equation (17).

References

Bertaut, E. F. (1955). *Acta Cryst.* **8**, 537.
Bertaut, E. F. (1956). *Acta Cryst.* **9**, 455.
Bertaut, E. F. (1960). *Acta Cryst.* **13**, 546, 643.
Cochran, W. (1955). *Acta Cryst.* **8**, 473.
Cochran, W. & Woolfson, M. M. (1955). *Acta Cryst.* **8**, 1.
Cramér, H. (1971). *Mathematical Methods of Statistics*, twelfth printing. Princeton Univ. Press.
Harker, D. & Kasper, J. S. (1948). *Acta Cryst.* **1**, 70.
Hauptman, H. (1971). *Z. Kristallogr.* **134**, 28.
Hauptman, H. & Karle, J. (1953*a*). *Solution of the Phase Problem.* ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.
Hauptman, H. & Karle, J. (1953*b*). *Acta Cryst.* **6**, 136.
Hauptman, H. & Karle, J. (1957). *Acta Cryst.* **10**, 267.
Hauptman, H. & Karle, J. (1958). *Acta Cryst.* **11**, 149.
Karle, I. L. (1970). *Acta Cryst.* B**26**, 1639.
Karle, J. (1971). *Acta Cryst.* B**27**, 2063.
Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181.
Karle, J. & Hauptman, H. (1953). *Acta Cryst.* **6**, 131.
Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635.
Karle, J. & Hauptman, H. (1957). *Acta Cryst.* 10, 515.
Karle, J. & Hauptman, H. (1958). *Acta Cryst.* **11**, 264.
Karle, J. & Karle, I. L. (1966). *Acta Cryst.* **21**, 849.
Klug, A. (1958). *Acta Cryst.* **11**, 515.
Naya, S., Nitta, I. & Oda, T. (1964). *Acta Cryst.* **17**, 421.
Naya, S., Nitta, I. & Oda, T. (1965). *Acta Cryst.* **19**, 734.
Tsoucaris, G. (1970). *Acta Cryst.* A**26**, 492.
Vaughan, P. A. (1958). *Acta Cryst.* **11**, 111.
Woolfson, M. M. (1954). *Acta Cryst.* **7**, 61.