

Exact Conditional Distribution of a Three-Phase Invariant in Space Group $P1$. IV. Further Improvements of Cochran-like Approximations

BY RICHARD I. SHRAGER, URI SHMUELI* AND GEORGE H. WEISS

Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20882, USA

(Received 4 October 1993; accepted 1 March 1994)

Abstract

A new approximate method of computing the conditional probability density function (c.p.d.f.) of a three-phase invariant is investigated, the results being compared to accurate calculations [Shmueli, Rabinovich & Weiss (1989). *Acta Cryst.* A45, 361–367]. A direct dependence on N , the number of atoms in the unit cell, is incorporated into the new approximation and its performance in the equal-atom case appears to be excellent over a large range of N values. The polynomial approximation of the recently published method [Posner, Shmueli & Weiss (1993). *Acta Cryst.* A49, 260–265] has been improved in both accuracy and range.

Introduction

An exact expression for the conditional probability density function (hereinafter c.p.d.f.) of a three-phase invariant in space group $P1$ has been derived and implemented in computer programs by Shmueli, Rabinovich & Weiss (1989*a,b*). The underlying joint probability density function (hereinafter j.p.d.f.) is hexavariate as the random variables of the problem are the real and imaginary parts of the three structure factors involved in the phase invariant. The computation of this exact c.p.d.f. requires the evaluation of a sixfold summation, the terms of which are products of Bessel-function series. This computation is therefore rather time consuming and cannot be proposed for routine application. Preliminary computations suggested that the exact c.p.d.f. of the three-phase invariant has a similar functional form to the approximate c.p.d.f. derived by Cochran (1955) on the basis of the central limit theorem, but the exact c.p.d.f.'s are more sharply peaked than the approximate ones. We have therefore investigated an approximation to the exact c.p.d.f. of the three-phase invariant that is expressed as a Cochran-type expres-

sion having the form

$$p(\Phi|\kappa') = [2\pi I_0(\kappa')]^{-1} \exp(\kappa' \cos \Phi), \quad (1)$$

where

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}} \quad (2)$$

is a three-phase invariant, κ' is a parameter derived from c.p.d.f.'s computed from the exact series (Shmueli *et al.*, 1989*a*) and $I_0(x)$ is the modified Bessel function of the first kind. We recall that Cochran's (1955) c.p.d.f. has the same form as (1) but with κ' replaced by

$$\kappa = 2N^{-1/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|, \quad (3)$$

where N is the number of atoms in the unit cell. The first attempt at this simple approximation was summarized by Posner, Shmueli & Weiss (1993). This work consisted of fitting a least-squares polynomial in $E_1 \equiv |E_{\mathbf{h}}|$, $E_2 \equiv |E_{\mathbf{k}}|$ and $E_3 \equiv |E_{-\mathbf{h}-\mathbf{k}}|$ to the ratio $\rho \equiv \kappa'/\kappa$, as well as expressing the ratio ρ as a product of a polynomial and an exponential function of the three structure-factor magnitudes and their powers. The numerical coefficients of these functions are based on extensive computations of exact c.p.d.f.'s for a range of E_i values and several discrete values of N . These coefficients have no direct dependence on N so that some sort of extrapolation is required beyond the range of N used in those calculations. It therefore seems desirable to devise a rapidly computable approximation, with the same conceptual basis as that given by Posner *et al.* (1993) but giving the user a more extended range of N and the possibility of treating the parameters E_1 , E_2 , E_3 and N on equal footings. A new algorithm for such an approximation is described in the next section. The following section contains some improvements in the polynomial fit already described by Posner *et al.* (1993) and the paper is concluded by a short comparison of the performance of the methods described. These methods illustrate the kind of modeling that may also be of use in approximating c.p.d.f.'s for space groups other than $P1$ whenever exact expressions for such c.p.d.f.'s become available.

* Permanent address: School of Chemistry, Tel Aviv University, 69 978 Tel Aviv, Israel.

Description of the new algorithm

The process of estimating approximate c.p.d.f.'s from (1) is based on exact c.p.d.f.'s computed for values of E_i , $i=1,2,3$, in the range 1.0 to 2.75 and initial values of N in the range 15 to 40. The purpose of the new algorithm, which also deals with the equal-atom case, is to generate a table of coefficients, yielding an accurate approximation to the c.p.d.f. of Φ . The table was generated as follows:

(1) As in the methods of Posner *et al.* (1993), for every set of arguments (E_1, E_2, E_3, N) not exceeding the upper limit of the E range investigated, the parameter κ' was computed which forced the Cochran-type c.p.d.f., given by (1), to agree with the exact c.p.d.f. at $\Phi = 0$.

(2) For every set of E 's (E_1, E_2, E_3) and all the values of N used with this set in exact computations, the ratios κ'/κ were fitted to a cubic form in $1/N$,

$$\rho = \kappa'/\kappa = 1 + (c_1/N) + (c_2/N^3), \quad (4)$$

where κ is given by (3). Equation (4) contains the dependence on N and is found to provide a good approximation for the c.p.d.f. of Φ in the range $N \geq 15$. It may be pointed out that the N^{-2} term was included in preliminary calculations but was later omitted because of the smallness of its coefficients.

(3) All the sets of E 's with a common sum were then grouped and the first four moments for each set of three E 's were computed. The first moment

$$m_1 = (E_1 + E_2 + E_3)/3 \quad (5)$$

is constant within each such group, while the higher moments

$$m_k = [(E_1 - m_1)^k + (E_2 - m_1)^k + (E_3 - m_1)^k]/3 \quad (6)$$

for $k = 2, 3, 4$ may of course be different for different sets with the same sum.

(4) Finally, within each group having a common sum of E 's, the number of triplets in such a group ranging from four to six, the coefficients c_1 and c_2 from (4) were fitted to linear combinations of the moments m_i :

$$c_1 = \sum_{k=1}^4 p_k m_k \quad (7)$$

$$c_2 = \sum_{k=1}^4 p_{k+4} m_k. \quad (8)$$

The coefficients p_i of these combinations form the table that is used for estimating the c.p.d.f. of Φ . The table contains nine columns, the first column having the sums of E 's on which the calculation is based and the $(i+1)$ th column containing p_i , with $i = 1, \dots, 8$. These data are collected in Table 1. The rows of zeros in Table 1 are to ensure that the behavior of the interpolation for small E 's will be reasonable, *i.e.* ρ will approach 1 as the E 's approach zero.

The estimation of the c.p.d.f. from the input values of E_1, E_2, E_3 and N proceeds in an essentially reverse order of steps to that outlined above. The steps that are required are:

(1) Sum the three E 's that were input, look up that sum in the first column of Table 1 and obtain the relevant p_i 's by linear interpolation on the sums (column 1) for $p_1 - p_4$ and linear interpolation on the cubes of the sums for $p_5 - p_8$.

(2) Compute the first four moments as indicated in (5) and (6) and evaluate c_1 and c_2 from (7) and (8).

(3) Compute κ from (3) and κ' from (4).

(4) Compute the approximate c.p.d.f. from (1).

The performance of this new approximation is discussed and illustrated in the final section.

An improved polynomial approximation

During the present study, some improvements were introduced into the computer program for exact c.p.d.f.'s, based on the algorithm of Shmueli *et al.* (1989a). These consisted mainly of enhancing the precision of the computation and precalculation of some trigonometric functions, which led to some reduction in the computing time. These improvements were followed by tests of the program and an extensive computation of exact c.p.d.f.'s of Φ was performed. Altogether, 1008 c.p.d.f.'s were computed, each in the range $0 \leq \Phi \leq 90$, comprising 84 different sets of E_1, E_2 and E_3 , the c.p.d.f. for each such set being computed for N varying from 15 to 70 in steps of 5. The E values in this computation ranged from 1.25 to 2.75 in steps of 0.25 and included the end-point of the range for all of the values of N considered. Specifically, E_1 ranged from 1.25 to 2.75, E_2 from E_1 to 2.75 and E_3 from E_2 to 2.75. For each of the exact c.p.d.f.'s, a parameter κ' was computed that constrained (1) to be equal to the exact c.p.d.f. at $\Phi = 0$ and 12 systems of equations,

$$\sum_{j=1}^5 a_{ij} q_j = \rho_i, \quad (9)$$

one for each value of N , were constructed and solved for the q_j 's by the least-squares method, where the index i ranges over all different c.p.d.f.'s with the same value of N , $\rho_i = \kappa'_i/\kappa_i$, κ_i being given by (3), depending on the appropriate values of E , and

$$a_{i1} = 1 \quad (10)$$

$$a_{i2} = E_1 E_2 E_3 \equiv T \quad (11)$$

$$a_{i3} = T(E_1 + E_2 + E_3) \quad (12)$$

$$a_{i4} = T(E_1^2 + E_2^2 + E_3^2) \quad (13)$$

$$a_{i5} = T(E_1 E_2 + E_2 E_3 + E_3 E_1). \quad (14)$$

The polynomial coefficients q_j , *i.e.* the least-squares solutions of (9) for the 12 different values of N , are

Table 1. Coefficients of the N -dependent combinations of moments from (7) and (8)

The table lists the p_j 's obtained from the fit described in the text. The variable Sum in the first column of the table corresponds to the sum of three E 's, which is taken as the lookup variable for the purpose of interpolation. The row of zeros at the top of each part of the table ensures that the output κ' will approach κ for small input E 's. That is, if Sum is less than 4.75, an interpolation between Sum = 0.00 and Sum = 4.75 applies as described in the text. The notation d + 02 means $\times 10^2$ etc.

Sum	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
0.00	0.d0	0.d0	0.d0	0.d0	0.d0	0.d0	0.d0	0.d0
4.75	2.9357431d+00	4.0872027d+00	-2.9851156d-01	-1.1196822d-01	9.1721603d+01	1.2617928d+03	-5.9033255d+02	6.2274343d+02
5.00	3.3248967d+00	4.3394953d+99	-3.1887020d-01	-1.8360029d-01	9.3291533d+01	1.4876488d+03	-7.2103437d+02	7.9259431d+02
5.25	3.6883663d+00	4.6308843d+00	-3.4575212d-01	-2.5946429d-01	1.0043987d+02	1.6948631d+03	-8.5539932d+02	9.8924909d+02
5.50	4.0169865d+00	4.8748876d+00	-3.3464682d-01	-3.5177149d-01	1.2712407d+02	1.9489354d+03	-1.0532391d+03	1.2458130d+03
5.75	4.3126518d+00	5.0850128d+00	-2.2799037d-01	-6.2947330d-01	1.8315531d+02	2.1993947d+03	-1.3528071d+03	1.7651114d+03
6.00	4.5915172d+00	5.1894809d+00	-8.5375540d-02	-1.0166698d+00	2.6215700d+02	2.5870258d+03	-1.6574841d+03	2.3939326d+03
6.25	4.8621795d+00	5.3384508d+00	1.3012670d-01	-1.7974944d+00	3.6597675d+02	2.9475226d+03	-2.0360387d+03	3.4997018d+03
6.50	5.1278799d+00	5.7346871d+00	1.6031388d-01	-3.5358131d+00	4.9757920d+02	3.0687927d+03	-2.1527290d+03	5.7425006d+03
6.75	5.4004982d+00	6.0339551d+00	-4.1840630d-01	-6.2138366d+00	6.4422968d+03	3.4465926d+03	-1.5729023d+03	8.9668723d+03
7.00	5.6774277d+00	5.3019869d+00	-5.5314723d-02	-6.4529352d+00	8.1348760d+02	5.0997264d+03	-2.1380393d+03	9.3239088d+03
7.25	5.9473356d+00	4.3058726d+00	3.0421661d-01	-6.5179276d+00	1.0193952d+03	7.1328239d+03	-2.6798146d+03	9.5733328d+03

given in Table 2. We can now recompute the i th c.p.d.f. from (1) by recovering κ'_i from the least-squares estimate of ρ_i and from the Cochran κ_i for these values of E_1, E_2, E_3 and N . This is done as follows:

(1) Compute the ratio from (9)–(14) and Table 2 for the required value of N , possibly by linear interpolation (if N is not a multiple of five).

(2) Compute κ' as $\rho\kappa$, where κ is given by (3).

(3) Compute the c.p.d.f. of Φ from (1).

For N between tabulated values, say $5i$ and $5(i+1)$, the ratio ρ_N is obtained to good accuracy by linear interpolation. Thus, for $5i < N < 5(i+1)$ and $3 < i < 14$, we have

$$\rho_N = \rho_{5i} + p(\rho_{5i+1} - \rho_{5i}), \quad (15)$$

where

$$p = (N - 5i)/5 \quad \text{or} \quad p = \left(\frac{1}{N} - \frac{1}{5i} \right) / \left[\frac{1}{5(i+1)} - \frac{1}{5i} \right]. \quad (16)$$

The two methods of linear interpolation lead to similar results. For $N > 70$, the extrapolation formula

$$\rho_N = 1 + (70/N)(\rho_{70} - 1) \quad (17)$$

may be used, where ρ_{70} is the value of ρ at $N = 70$.

Comparative computations

Several computation runs were carried out with the approximations described in the previous sections and discrepancies between the c.p.d.f.'s so obtained and those computed from the exact expressions of Shmueli *et al.* (1989a) were expressed in terms of R factors defined as

$$R = \left\{ \frac{\sum_{i=1}^{n_\Phi} (p'_{\text{ex}} - p'_{\text{approx}})^2}{\sum_{i=1}^{n_\Phi} [p'_{\text{ex}}]^2} \right\}^{1/2} \quad (18)$$

Table 2. Least-squares estimates of the polynomial coefficients in (9)

Each row of the table contains the value of N , the number of atoms in the unit cell and coefficients of the polynomial defined in (9) to (14) that correspond to this N .

N	q_1	q_2	q_3	q_4	q_5
15	0.427910	1.134285	-0.311552	0.125963	-0.052796
20	0.883948	0.286319	-0.065698	0.040241	-0.024926
25	0.900867	0.224664	-0.047014	0.023851	-0.013966
30	0.910243	0.191166	-0.038569	0.016973	-0.009338
35	0.918940	0.165820	-0.032786	0.013165	-0.006908
40	0.927380	0.144764	-0.028180	0.010718	-0.005473
45	0.935022	0.127331	-0.024456	0.009007	-0.004539
50	0.940305	0.115162	-0.021989	0.007789	-0.003822
55	0.944443	0.105556	-0.020064	0.006895	-0.003317
60	0.951988	0.091054	-0.016835	0.006101	-0.003162
65	0.962720	0.072795	-0.012722	0.005209	-0.003046
70	0.969729	0.062769	-0.010818	0.003997	-0.002123

Table 3. Exact versus approximate c.p.d.f.'s in the range (15–70)

The table lists the R factors (see text) that measure the discrepancy between the exact c.p.d.f. of Φ and the Cochran (1955) c.p.d.f. ($R_{\text{ex-Coch}}$), the c.p.d.f. based on the data in Table 1 ($R_{\text{ex-SSW}}$) and the c.p.d.f. based on the data in Table 2 ($R_{\text{ex-PSW}}$). The E values in this computation are: $E_1 = 2.09, E_2 = 2.35$ and $E_3 = 2.63$.

N	$R_{\text{ex-Coch}}$	$R_{\text{ex-SSW}}$	$R_{\text{ex-PSW}}$
15	0.3906	0.0368	0.0338
20	0.2864	0.0183	0.0187
25	0.2285	0.0130	0.0136
30	0.1912	0.0102	0.0107
35	0.1650	0.0084	0.0089
40	0.1456	0.0072	0.0076
45	0.1305	0.0063	0.0066
50	0.1184	0.0055	0.0058
55	0.1084	0.0049	0.0052
60	0.1002	0.0044	0.0047
65	0.0935	0.0040	0.0041
70	0.0877	0.0033	0.0031

where n_Φ is the number of Φ values at which the c.p.d.f.'s were computed, p'_{ex} is the value of the exact c.p.d.f. at the i th point and p'_{approx} is the value of an

approximate c.p.d.f. of Φ at the i th point. In most computations, $n_\phi = 19$ and Φ was varied from 0 to 90° in steps of 5°. The approximations considered are: (i) Cochran's (1955) c.p.d.f. of Φ , based on the central limit theorem; (ii) the c.p.d.f. of Φ based on the new algorithm described in this paper; and (iii) the c.p.d.f. of Φ based on an improved polynomial approximation based on the method of Posner *et al.* (1993).

The agreement of approximations (ii) and (iii) with the exact c.p.d.f. is very good throughout the range $15 \leq N \leq 70$, as can be seen from Table 3. The discrepancy of Cochran's (1955) c.p.d.f. and the exact one is quite considerable in this range of N ,

although it decreases slowly with increasing N . Both new approximations have a similar behaviour: the value of R is highest at the low end of the N range and decreases with increasing N ; their performance is similar and, in general, very good.

References

- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.
 POSNER, Y., SHMUELI, U. & WEISS, G. H. (1993). *Acta Cryst.* **A49**, 260–265.
 SHMUELI, U., RABINOVICH, S. & WEISS, G. H. (1989a). *Acta Cryst.* **A45**, 361–367.
 SHMUELI, U., RABINOVICH, S. & WEISS, G. H. (1989b). *Acta Cryst.* **A45**, 367–371.

Acta Cryst. (1994). **A50**, 636–643

Rotation of Real Spherical Harmonics

BY ZHENGWEI SU AND PHILIP COPPENS

Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14214-3094, USA

(Received 7 January 1994; accepted 18 March 1994)

Abstract

Formulae for the rotation of real spherical harmonic functions are presented. To facilitate their application, values of the matrices $d_{m'm}^{(l)}(\pi/2)$, which occur in the equations, are tabulated for $1 \leq l \leq 8$ and $0 \leq m', m \leq l$. The application of the equations to spherical harmonic functions with normalization commonly used in charge-density analysis is described.

Introduction

The real spherical harmonic functions are extensively used for the description of atomic orbitals and as density basis functions in the analysis of experimental charge densities. In order to recognize the local or global symmetry of a particular site, it is often necessary to rotate the coordinate system after completion of a theoretical calculation or an experimental charge-density analysis. In the multipole analysis of charge densities, for example, application of local symmetry constraints requires the use of a local coordinate system on each of the atoms (Hansen & Coppens, 1978). For subsequent calculation of molecular properties, such as molecular electrostatic moments, it is necessary to rotate the functions to a common coordinate system.

The treatment given starts with the equations by Steinborn & Ruedenberg (1973) for the rotation of complex spherical harmonic functions and is similar to that described earlier by Cromer, Larson & Stewart (1976); however, expressions are given for both unnormalized and normalized spherical harmonic functions, the latter with normalization appropriate for either wave functions or density functions. Explicit numerical values are given for the matrices (up to $l = 8$) that occur in the equations, thus facilitating their application. In addition, a number of inadvertent errors in the earlier publication have been eliminated.

Coordinate-system rotations

Let (r, θ, φ) and (r, θ', φ') be the spherical coordinates of a vector

$$\mathbf{X} = (x_1, x_2, x_3) \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{pmatrix} = (x'_1, x'_2, x'_3) \begin{pmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \mathbf{e}'_3 \end{pmatrix}.$$

The unitary matrix that transforms the two right-handed Cartesian bases \mathbf{e} and \mathbf{e}' can be written in terms of Eulerian angles α , β and γ (Arfken, 1970; Edmonds, 1974; Steinborn & Ruedenberg, 1973),