The use of the reflection *ugu* illustrates this difference. The unrestricted phase φ_{01u} may be set to some arbitrary value since it serves only to fix the origin on a polar *y* axis, but the phase φ_{ugu} (or for that matter any other unrestricted phase) must be assigned a value which conforms with the four permissible origins along (x,z) [*i.e.* $(0,0)$, $(\frac{1}{2},0)$, $(0,\frac{1}{2})$ and $(\frac{1}{2},\frac{1}{2})$], and the origin on the *y* axis fixed by $01u$.

The problem is that the value of φ_{ugu} is unknown at the start of a direct-methods process, and it is necessary to try a series of values between 0 and 2π. Typically, φ_{ugu} could be initiated at $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$ in four separate calculations and be permitted to vary in the final cycles of phase refinement. An interesting aspect of this procedure is that if the actual value of φ_{ugu} is significantly different from 0 and π, it will also serve to specify the enantiomorph, and the multi-solution process need only vary its value between 0 and π. However, if the value of φ_{ugu} happens to be close to 0 or π then another reflection must be specified to fix the enantiomorph. For this reason it is strongly recommended that an additional phase is specified for independent enantiomorphic discrimination.

(i) *Test uuu* (class 13, Table 3):

$$\mathbf{n}' = (1,u,1)\,\mathbf{U}^{-1} = (u,u,1)$$

and

$$\varphi' = (u,u,1)\,(0,0,q) = q.$$

Since φ_{uuu} is unrestricted it may be used to specify the enantiomorph provided it is significantly different from q and $q+\pi$.

(ii) *Test* $0uu$ (class 3, Table 3):

$$\mathbf{n}' = (0,u,1)\,\mathbf{U}^{-1} = (u,g,0)$$

and

$$\varphi' = (u,g,0)\,(0,0,q) = 0.$$

Since φ_{0uu} is unrestricted, it may be used to specify the enantiomorph provided it has a value significantly different from 0 and π.

References

HALL, S. R. (1982). *GENTAN:* Application of a General Tangent Phasing Procedure. *XTAL* System of Crystallographic Programs. Univ. of Maryland.

HAUPTMAN, H. & KARLE, J. (1954). *Solution of the Phase Problem* I: *The Centrosymmetric Crystal. ACA Monogr.* No. 3. Pittsburgh: Polycrystal Book Service.

HAUPTMAN, H. & KARLE, J. (1956). *Acta Cryst.* **9**, 45–55.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.

RODGERS, D. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER. New York: Plenum Press.

Acta Cryst. (1983). A**39**, 26–28

The Probability of Large Structure Amplitudes: The Space Group $P\bar{1}$

BY A. J. C. WILSON*

Department of Physics, The University of Birmingham, Birmingham B15 2TT, England

(*Received* 14 *May* 1982; *accepted* 6 *August* 1982)

Abstract

Application of limit theorems valid for large values of the sum of independent random variables shows that for an equiatomic structure with n atoms in the asymmetric unit in the space group $P\bar{1}$ the probability density distribution of the structure amplitude F for $|F|$ large is

$$p(x)\,\mathrm{d}x = [n/2\pi(1-x^2)]^{1/2}\,(e/\pi)^{n/2}\,(1-x)^{(n-1)/2}\,\mathrm{d}x$$

where $x = F/2nf$ is the unitary structure amplitude and f is the atomic scattering factor. The expression will be somewhat different for other space groups.

1. Introduction

The probability distribution of structure amplitudes is a special case of the random-walk problem. Expressions valid for resultants small compared with the maximum possible are readily available, but the standard sources do not give expressions valid for large resultants (Wilson, 1980). The present paper uses a limit theorem other than the central-limit theorem to derive an approximate distribution for large structure amplitudes in the space group $P\bar{1}$.

* Address from 1 January 1983: University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England.

0567-7394/83/010026-03$01.50

© 1983 International Union of Crystallography

The structure amplitude F for an approximately equiatomic structure in space group $P\bar{1}$ is given by

$$F/2f \equiv y = \cos\theta_1 + \cos\theta_2 + \cdots + \cos\theta_n, \quad (1)$$

where f is the atomic scattering factor and θ_i is the phase angle of the ith atom in the asymmetric unit of n atoms. If θ_i is uniformly distributed in the range 0 to 2π, the probability of $\cos\theta_i$ having a particular value is given by the well-known probability distribution

$$p(X_i)\,\mathrm{d}X_i = \pi^{-1}(1-X_i^2)^{-1/2}, \quad |X_i| \le 1,$$
$$= 0, \text{ otherwise.} \quad (2)$$

The maximum value of $|y|$ is obviously n, though its typical values are much smaller, of the order of $n^{1/2}$. For these typical values the usual central-limit theorem gives the familiar centric distribution function (Wilson, 1949). There are, however, other limit theorems, apparently first given by Cramér (1938), expressing the probability of resultants like

$$y = X_1 + X_2 + \cdots + X_n \quad (3)$$

in terms of the moment-generating function of the variables X_i, assumed to be independent and to have the same probability distribution, but without making the assumption that y is small compared with $n|X|_{\max}$. A simple statement of some of the theorems is given by Petrov (1965), whose notation is used here as far as possible.

The moment-generating function is defined by

$$R(h) \equiv \langle \exp(hX) \rangle = \int \exp(hX)\, p(X)\, \mathrm{d}X. \quad (4)$$

Other functions needed are its logarithm, the cumulant-generating function,

$$K(h) \equiv \log_e R(h), \quad (5)$$

the derivative of the cumulant-generating function,

$$m(h) \equiv K'(h) = \langle X \exp(hX) \rangle / \langle \exp(hX) \rangle, \quad (6)$$

and the second derivative

$$\sigma^2(h) \equiv K''(h) = \frac{\langle X^2 \exp(hX) \rangle}{\langle \exp(hX) \rangle} - \frac{\langle X \exp(hX) \rangle^2}{\langle \exp(hX) \rangle^2}. \quad (7)$$

The notation in (6) and (7) reflects the fact that $m(0)$ and $\sigma^2(0)$ are the ordinary mean and variance of X. The theorems in question state that, under various alternative conditions that appear to be satisfied for the present application, the probability of y attaining a value equal to or greater than a fraction x of its maximum value n is given by

$$P(y \ge nx) = \frac{\exp\{nK(h) - nhx\}}{(2\pi n)^{1/2}\, h\sigma(h)} \{1 + o(1)\}, \quad (8)$$

where in equation (8) h is the unique real root of the equation

$$m(h) = x, \quad (9)$$

and $o(1)$ indicates a function of n that is negligible in comparison with unity when n is sufficiently large.

For reasonable distributions the complementary cumulative distribution function given by (8) can be differentiated to give the more familiar probability density distribution $p(x)$. The relation is

$$1 - P = \int_{-\infty}^{x} p(x)\, \mathrm{d}x, \quad (10)$$

so that

$$p(x)\, \mathrm{d}x = -(\mathrm{d}P/\mathrm{d}x)\, \mathrm{d}x, \quad (11)$$

$$p(x)\,\mathrm{d}x = P(y \ge nx) \times \{-nm(h)\,h' + nh + nh'\,x + h^{-1}\,h' + \sigma^{-1}\,\sigma' + \cdots\}\,\mathrm{d}x, \quad (12)$$

where the primes denote differentiation with respect to x. Since h has been chosen to fulfil (9), the first and third terms cancel each other. Since the fourth and fifth do not contain n, they can be absorbed into the $o(1)$, but it will appear later that $h\sigma(h)$ is almost constant for large x, and they would in any case practically cancel each other. The probability density distribution is thus

$$p(x)\,\mathrm{d}x = (n/2\pi\sigma^2)^{1/2} \exp\{nK(h) - nhx\} \times \{1 + o(1)\}\,\mathrm{d}x \quad (13)$$
$$= (n/2\pi\sigma^2)^{1/2} R^n(h) \exp(-nhx)\{1 + o(1)\}\,\mathrm{d}x. \quad (14)$$

2. Space group $P\bar{1}$

2.1. *General expressions*

The expressions just given apply for any moment-generating function. For the structure amplitudes in the space group $P\bar{1}$ the moment-generating function is

$$R(h) = \frac{1}{\pi} \int_{-1}^{1} (1 - X^2)^{-1/2} \exp(hX)\, \mathrm{d}X \quad (15)$$
$$= I_0(h), \quad (16)$$

where I_0 is the zero-order modified Bessel function of the first kind (Abramowitz & Stegun, 1964, formula 9.6.18). The remaining functions become

$$K(h) = \log_e I_0(h) \quad (17)$$
$$m(h) = I_0'(h)/I_0(h) = I_1(h)/I_0(h) \quad (18)$$
$$= x \quad (19)$$
$$\sigma^2(h) = I_0''(h)/I_0(h) - [I_0'(h)/I_0(h)]^2 \quad (20)$$
$$= 1 - x/h - x^2. \quad (21)$$

Examination of tables and series expansions of Bessel functions shows that for small h the function $m(h)$ increases as $\frac{1}{2}h$, then bends over, and for large h it approaches unity as $1 - (2h)^{-1}$. Equation (9) thus has a single root for all values of x between 0 and 1, as expected, which is given approximately by

$$h = 1/2(1-x) + \tfrac{1}{4} + \cdots \qquad (22)$$

for large h (x approaching unity); σ^2 is thus approximately $\frac{1}{2} - x^2$ for h and x small and $(1-x)^2(1+x)$ for x approaching unity.

2.2. *Small resultants*

It is reassuring that, with the values of h and σ just derived for small x, (13) reduces to the centric distribution. For small h the cumulant-generating function is

$$K(h) = \log_e I_0(h) = \tfrac{1}{4}h^2 + \ldots, \qquad (23)$$

so that (13) becomes

$$p(x)\,\mathrm{d}x = (n/\pi)^{1/2} \exp\{\tfrac{1}{4}n(2x)^2 - 2nx^2\}\,(1+\ldots)\,\mathrm{d}x$$
$$= (n/\pi)^{1/2} \exp(-nx^2)\,\mathrm{d}x + \ldots. \qquad (24)$$

The normalized structure amplitude is related to x by

$$E = (2n)^{1/2}x, \qquad (25)$$

so that (24) becomes

$$p(E)\,\mathrm{d}E = (2\pi)^{-1/2} \exp(-\tfrac{1}{2}E^2)\,\mathrm{d}E + \ldots, \qquad (26)$$

the correct centric expression when E is regarded as taking on negative as well as positive values.

2.3. *Large resultants*

For x approaching unity, and hence h becoming large, the Bessel functions may be replaced by their asymptotic expansions (Abramowitz & Stegun, 1964, formula 9.7.1). The cumulant-generating function becomes

$$K(h) = \log_e I_0(h) = h - \tfrac{1}{2}\log_e(2\pi) - \tfrac{1}{2}\log_e h + \ldots \qquad (27)$$

and

$$m(h) = 1 - (2h)^{-1} + \ldots \qquad (28)$$

$$K(h) - hm(h) = \tfrac{1}{2} - \tfrac{1}{2}\log_e(2\pi) - \tfrac{1}{2}\log_e h + \ldots$$
$$= \tfrac{1}{2}\log_e[e(1-x)^{(n/2)} - 1\,\mathrm{d}x. \qquad (29)$$

on using (22). Equation (13), with the value of σ^2 following (22), now gives

$$p(x)\,\mathrm{d}x = [n/2\pi(1+x)]^{1/2}(e/\pi)^{n/2}(1-x)^{(n/2)-1}\,\mathrm{d}x. \qquad (30)$$

This looks reasonable for the moderate and large values of x and n to which it is expected to apply, since it diminishes rapidly as x increases and vanishes for $x = 1$.

Equation (30) cannot be expected to apply for n small, but it is nevertheless interesting to consider the special cases $n = 1$ and $n = 2$. For $n = 1$

$$p(x)\,\mathrm{d}x = (e/2)^{1/2}\,\pi^{-1}(1-x^2)^{-1/2}\,\mathrm{d}x, \qquad (31)$$

which is its correct value, as given by (2), except for the factor $(e/2)^{1/2} = 1{\cdot}1658\ldots$. It is not clear whether this close reproduction of the original distribution, after all the intervening approximations, is anything more than a curiosity. For $n = 2$

$$p(x)\,\mathrm{d}x = (e/\pi)\,[\pi(1+x)]^{-1/2}\,\mathrm{d}x. \qquad (32)$$

This is qualitatively similar to the correct distribution, as shown in Fig. 3.1 of Srinivasan & Parthasarathy (1976), but shows less variation with x.

3. Extension to other space groups

With some adjustments of factors of two, the expressions derived above would apply to the real and imaginary parts of structure amplitudes in the space group $P1$. There is no obvious way, however, of passing to the distribution of $|F|$ for this space group, since the real and the imaginary parts are highly correlated when large. The moments and cumulants, up to about the tenth, are known for the higher space groups (Wilson, 1978; Shmueli & Wilson, 1981; Shmueli, 1982; and papers in preparation), but the cumulant-generating function expressed in ascending powers would have to be inverted into an asymptotic expansion in some way before the equivalent of (27)–(29) could be obtained. Further work may be worth while, as the Edgeworth or Gram–Charlier expansions developed by Shmueli and Wilson show unacceptable negative probabilities for some ranges of structure amplitude greater than about $2\Sigma^{1/2}$. Some progress has been made with the space groups of the point group $2/m$.

I am indebted to Mr D. M. Grove of the Department of Statistics for checking some of the manipulations.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1964). *Handbook of Mathematical Functions*. Washington: US Government Printing Office.

CRAMÉR, H. (1938). *Sur un Nouveau Théorème–Limite de la Théorie des Probabilités*. No. 736 in the series *Actualités Scientifiques et Industrielles*. Paris: Hermann.

PETROV, V. V. (1965). *Theory Probab. Its Appl.* (*USSR*), **10**, 287–298.

SHMUELI, U. (1982). *Acta Cryst.* A**38**, 362–371.

SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 342–353.

SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 381–321.

WILSON, A. J. C. (1978). *Acta Cryst.* A**34**, 986–994.

WILSON, A. J. C. (1980). *Technometrics*, **22**, 629–630.