Acta Cryst. (1977). A33, 838–844

Quantitative Statistical Criteria for Use in the Cumulative, Semi-cumulative and Distribution Function Tests for Centrosymmetry

BY M. N. PONNUSWAMY AND S. PARTHASARATHY

Department of Crystallography & Biophysics, University of Madras, Guindy Campus, Madras*-600025, *India*

(*Received* 23 *February* 1977; *accepted* 9 *April* 1977)

Several criteria suitable for taking a decision regarding the centrosymmetry of crystals by the study of the goodness-of-fit of the observed and expected cumulative, semi-cumulative and distribution functions of normalized structure amplitudes are suggested. Practical aspects of the various tests are discussed. From a study of the performance of these tests on four crystals a safe method of conducting these tests is suggested.

1. Introduction

The cumulative function test (Howells, Phillips & Rogers, 1950) and the distribution function test (Ramachandran & Srinivasan, 1959) are two of the earliest tests for centrosymmetry† (see Srinivasan & Parthasarathy, 1976, for a recent summary). The semi-cumulative function test (Hargreaves & Gogoi, 1966; Srikrishnan & Parthasarathy, 1970) was later proposed. The power of these tests has not so far been fully realized since no quantitative criteria for the goodness-of-fit of the observed and expected theoretical distributions are utilised. Several criteria for goodness-of-fit are available in the theory of statistics and this paper aims at showing how these criteria could be adapted to make the above three tests for centrosymmetry quantitative.

Some of the goodness-of-fit tests require actual values of the normalized structure amplitudes (denoted by y) of *all* the reflexions in a given region of $\sin\theta/\lambda$. Since, in practice, such complete data are unavailable because some reflexions are too weak to be observed, it would be useful to develop tests which could be conducted with only the reflexions whose normalized intensities are greater than the threshold value, say y_t. The modified distributions of y needed for such a case are derived in § 2 and used in conducting the various goodness-of-fit tests.‡

To obtain the best results from these tests, the normalized intensities need to be sufficiently accurate (Rogers, Stanley & Wilson, 1955). Further, for conducting these tests with computers it would be useful to obtain an analytic expression¶ for $\langle I_o \rangle$ as a function of $\sin\theta/\lambda$ (hereafter s). A procedure for doing this is described in § 5.

We shall consider six different goodness-of-fit tests, namely (i) Pearson's chi-square test, (ii) likelihood ratio test, (iii) binomial test, (iv) Kolmogorov test, (v) Smirnov–Cramer–Von Mises test and (vi) a Neyman–Barton smooth test. Of these (i) and (ii) provide quantitative criteria for goodness-of-fit for the distribution function. The statistic in (iii) provides a criterion for making the semi-cumulative function test quantitative. Tests (iv)–(vi) provide quantitative criteria for goodness-of-fit for the cumulative function. Of these (i)–(iii) require ordering of data into histogram bins and the others require unbinned data. In statistical theory these tests come under the domain of hypothesis testing, described in standard books on statistics (Eadie, Drijard, James, Roos & Sadoulet, 1971; Siegel, 1956; – hereafter abbreviated as E, 1971; S, 1956 respectively). To make the present discussion clear we shall summarize the methodology in § 3. The details of each of the tests are described in § 4. In order that the performance of these tests could be studied, they were tried on the observed data of four crystals, and the results are discussed in § 5.

In this paper we shall deal with tests for centrosymmetry for crystals which satisfy the requirements of the basic Wilson (1949) distributions. The modifications needed for conducting these tests in crystals with heavy atoms or pseudosymmetry are briefly indicated in § 6. It is convenient to use abbreviations *C* and *NC* to denote the terms *centrosymmetric* and *non-centrosymmetric* respectively.

2. Probability distribution functions for normalized intensities and normalized structure amplitudes valid for truncated data

Let $P(x)$ be the probability density function (abbreviated as pdf), and $N(x)$ the cumulative function of a random variable x, $0 \le x < \infty$. Let $P_t(x)$ denote the normalized pdf of x when the range of x is restricted to $x_t \le x < \infty$ (*i.e.* when x is truncated). $P_t(x)$ is related

* Contribution No. 453.

† Since the different statistical tests are affected differently by any given type of error in the observed intensities it is preferable to conduct different tests in order to arrive at a decision (Rogers, Stanley & Wilson, 1955).

‡ The binomial test is an exception [for details see § 4(iii)].

¶ $\langle I_o \rangle$ is needed to convert the observed intensities I_o into normalized intensities z through the relation $z = I_o/\langle I_o \rangle$. Note that $z = y^2$, where y is the normalized structure amplitude.

to $P(x)$ through

$$P_t(x)=\frac{P(x)}{\int_{x_t}^{\infty}P(x)\mathrm{d}x}=\frac{P(x)}{1-N(x_t)},\quad x_t\le x<\infty\,. \quad (1)$$

If the Wilson (1949) distribution functions are applied to (1) the pdf of z applicable to the truncated data, $z_t\le z<\infty$, is

$$P_t(z)=\exp(-z)/\exp(-z_t)\text{ for }NC$$
$$=(2\pi z)^{-1/2}\exp(-z/2)/\mathrm{erfc}\,(z_t/2)^{1/2}\text{ for }C\,. \quad (2)$$

The corresponding pdf of the normalized structure amplitude y valid for $y_t\le y<\infty$ is

$$P_t(y)=2y\exp(-y^2)/\exp(-y_t^2)\text{ for }NC$$
$$=\left(\frac{2}{\pi}\right)^{1/2}\exp(-y^2/2)/\mathrm{erfc}\,(y_t/\sqrt{2})\text{ for }C\,. \quad (3)$$

Evidently,

$$z_t=y_t^2\,. \quad (4)$$

When $z_t=y_t=0$, (3) and (4) reduce to the standard results valid for the untruncated data.

3. Methodology of conducting the various tests

The methodology of conducting these tests may be summarized in the following steps (for more details see *E*, 1971; *S*, 1956). (i) State the null hypothesis H_0. (ii) Define a suitable test statistic, say S, as a goodness-of-fit criterion and obtain its sampling distribution under H_0, say $P(S)$. Often it is simpler to use a sufficiently large sample for the asymptotic form of $P(S)$ to be obtained easily. (iii) Specify a level of significance α. In statistical applications the usual value of α is 0·05 (though values such as 0·01 and 0·1 are occasionally used). (iv) For this value of α obtain the critical value (say S_c) of the statistic from the sampling distribution $P(S)$. This partitions the range of values of S into two regions called the region of acceptance, say A, and the region of rejection, say R (also called the critical region). (v) From the sample data compute the observed value of the statistic as S_o. (vi) Take a decision which consists in accepting H_0 (at the chosen level α) if S_o lies in the region A and in rejecting H_0 if S_o lies in the critical region R.

Unless stated otherwise, we shall take the null hypothesis H_0 to be that the crystal is C. Equivalently we can state H_0 as*

$$H_0: P_t(y)=\left(\frac{2}{\pi}\right)^{1/2}\exp(-y^2/2)/\mathrm{erfc}\,(y_t/\sqrt{2}),$$
$$y_t\le y<\infty\,. \quad (5)$$

4. Description of the various tests

(i) *Pearson's chi-square test*

Suppose that the n observations are ordered into k histogram bins defined by the set of k intervals for the normalized structure amplitudes $y: y_{i-1}\le y<y_i$, $i=1,2,\ldots,k$. Let n_i be the number of reflexions in bin i. Let p_i be the probability content of bin i under H_0. Thus, if $P_o(y)$ is the pdf of y under H_0, then

$$P_i=\int_{y_{i-1}}^{y_i}P_o(y)\mathrm{d}y,\quad i=1,2,\ldots,k\,, \quad (6)$$

where $P_o(y)$ is the function $P_t(y)$ given in (5). The statistic in this case can be shown to be (*E*, 1971).

$$T=\frac{1}{n}\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{p_i}\,. \quad (7)$$

The sampling distribution of T under H_0 follows the chi-square distribution with $k-1$ degrees of freedom [denoted hereafter as $\chi^2(k-1)$]. The critical values of χ^2 are tabulated (*S*, 1956) for different degrees of freedom k and for different values of α. Hence the critical value of T required for conducting this test for any given α can be read from a χ^2 table.

Note the following practical points regarding this test: (*a*) The asymptotic $\chi^2(k-1)$ distribution of (7) requires that the number of events expected under H_0 be greater than five. According to Cochran (1952, 1954) this criterion can be relaxed to the level that not more than 20% of the bins have expectations less than five. (*b*) Since too few bins carry too little information and since too many bins may lead to too few events per bin, it would be useful to obtain the optimum number of bins k into which a given set of n reflexions can be grouped. An expression from which the optimum value of k could be obtained is given in *E* (1971). (*c*) For a given number (k) of bins, the bins should be chosen so as to have equal probability content under H_0.

(ii) *Likelihood-ratio test*

With the notation used in § 4(i) the test statistic for this case can be written as (*E*, 1971)

$$\Lambda=-2\log_e\left[n^n\prod_{i=1}^{k}\left(-\frac{p_i}{n_i}\right)^{n_i}\right]. \quad (8)$$

Λ behaves asymptotically as $\chi^2(k-1)$.

(iii) *Binomial test*

It is known that the $P(y)$ curves for the centric and acentric Wilson distributions intersect at the points $y=0{\cdot}439$ and $y=1{\cdot}704$ (Srikrishnan & Parthasarathy, 1970). These points divide the range $0\le y<\infty$ into three intervals, namely, $0\le y<0{\cdot}439$, $0{\cdot}439\le y<1{\cdot}704$

---

* In the binomial test we take H_0 as

$$P(y)=\left(\frac{2}{\pi}\right)^{1/2}\exp(-y^2/2),\ 0\le y<\infty$$

since no truncation of data is needed.

and $1{\cdot}704 \leq y < \infty$ (called intervals 1, 2 and 3 respectively). The reflexions belonging to interval 2 can be conveniently characterized as medium while those belonging to classes 1 and 3 as weak and strong respectively. It is convenient to name the reflexions belonging to interval 2 as class M and those belonging to intervals 1 and 3 as class $\overline{M}$. It is known that the area under the $P(y)$ curve and between the lines $y=0{\cdot}439$ and $y=1{\cdot}704$ is 0·571 for the C case and 0·768 for the NC case. The null hypothesis that the given crystal is C can be equivalently stated as

$$H_0: p = 0{\cdot}571\,, \tag{9}$$

where p is the proportion of reflexions belonging to class M. Let us denote $1-p$ by q, the proportion of reflexions belonging to class $\overline{M}$. The alternative hypothesis H_1 may be stated as*

$$H_1: p > 0{\cdot}571\,. \tag{10}$$

The test statistic is the number x of reflexions (out of a total of n) that belong to class M. The sampling distribution of x is the binomial distribution given by

$$P(x) = \binom{n}{x} p^x q^{n-x}\,. \tag{11}$$

If $p \simeq 0{\cdot}5$ and $n > 25$ the sampling distribution of x is approximately normal with mean np and variance npq. Hence H_0 may be tested by the statistic B defined by

$$B = \frac{x \pm 0{\cdot}5 - np}{(npq)^{1/2}}\,. \tag{12}$$

The sampling distribution of B is the standard normal distribution with zero mean and unit variance. In (12), $x+0{\cdot}5$ is to be used when $x<np$ and $x-0{\cdot}5$ when $x>np$ (this is called the correction for continuity). The critical value of B for the one-tailed test (for any α) can therefore be read from a table of the standard normal distribution.

* This is because in the present case the alternative is that the crystal is NC in which case $p=0{\cdot}768$ which is $>0{\cdot}571$.

(iv) *Kolmogorov test*

Suppose the set of n observed values of normalized structure amplitudes is ordered such that

$$y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}\,. \tag{13}$$

From (13) the observed cumulative function $N_n(y)$ is defined by

$$N_n(y) = \begin{cases} 0 & \text{if } y \leq y_{(1)} \\ i/n & \text{if } y_{(i)} \leq y < y_{(i+1)},\ i = 1,2,\ldots,n-1 \\ 1 & \text{if } y_{(n)} \leq y\,. \end{cases} \tag{14}$$

Let the cumulative function of y under H_0 be denoted by $N_0(y)$. From (5) we can show that, for the data with a y cut-off, H_0 can be written as

$$H_0: N_0(y) = [\mathrm{erf}\,(y/\sqrt{2}) - \mathrm{erf}\,(y_t/\sqrt{2})]/\mathrm{erfc}\,(y_t/\sqrt{2})\,, \quad y_t \leq y < \infty\,. \tag{15}$$

The test statistic D_n is defined to be the maximum deviation of the observed cumulative function from that expected under H_0, *i.e.*

$$D_n = \max |N_n(y) - N_0(y)|\,. \tag{16}$$

From the asymptotic sampling distribution of D_n (valid for $n \geq 80$) the critical values of $(n)^{1/2}D_n$ have been shown to be 1·63, 1·36 and 1·22 respectively corresponding to $\alpha = 0{\cdot}01$, 0·05 and 0·1 (E, 1971).

(v) *Smirnov–Cramer–Von Mises test*

The statistic W^2 in this case is the expected mean square difference between the observed cumulative function and that under H_0, namely $N_0(y)$. It is shown that (E, 1971)

$$W^2 = \frac{1}{n}\left[\frac{1}{12n} + \sum_{i=1}^{n}\left\{N_0(y_{(i)}) - \left(\frac{2i-1}{2n}\right)\right\}^2\right]. \tag{17}$$

From the asymptotic characteristic function of nW^2 (valid for $n \geq 3$) the critical values of nW^2 corresponding to the level of significance $\alpha = 0{\cdot}01$, 0·05 and 0·1 have been shown to be 0·743, 0·461 and 0·347 respectively.

Table 1. *Details of the crystal structures on which the various statistical tests were conducted*

Crystal*	Space group	Name of the crystal	Molecular formula	$\langle z \rangle$	No. of reflexions in Set B	No. of reflexions in Set C
I	$P\bar{1}$	Dimethyl ester of *meso*-tartaric acid	$C_6H_{16}O_6$	1·009	828	697
II	$P\bar{1}$	5α-Hydroxy-6α-4′[5′-methylpyrimidin-2′-one] dihydrothymine	$C_{10}H_{12}N_4O_4.H_2O$	0·996	1201	922
III	$P1$	L-*N*-Acetylhistidine monohydrate	$C_8H_{11}N_3O_3.H_2O$	1·008	1053	1034
IV	$P1$	11,11-Dimethyltricyclo [4,4,1,$0^{1,6}$]undeca-2,4,7,9-tetrane	$C_{13}H_{14}$	0·982	1039	1010

* The structures I to IV were solved by Kroon & Kanters (1973); Karle (1969); Kistenmacher, Hunt & Marsh (1972); Bianchi, Morosi, Mugnoli & Simonetta (1973) respectively.

(vi) *Neyman–Barton smooth test*

Obtain from the ordered set (13) a second ordered set x_i by the transformation

$$x_i = N_0(y_{(i)}), \quad i=1,2,\ldots,n, \tag{18}$$

where $N_0(y)$ is the cumulative function under H_0 (see equation 15). Under H_0 the x_i are uniformly distributed in the interval $0 \le x < 1$. The statistics in this test are constructed for specified departures from uniformity. The following two statistics could be used in practice:

$$p_s^2 = \frac{1}{n}\sum_{r=1}^{s}\left[\sum_{i=1}^{n} l_r(x_i)\right]^2, \quad s=1,2, \tag{19}$$

where the functions $l_r(x)$ are the Legendre polynomials of order r, orthonormal on (0,1). If $n \ge 20$, it is shown that the sampling distributions of the statistics p_1^2 and p_2^2 (see equation 19) are $\chi^2(1)$ and $\chi^2(2)$ respectively (E, 1971).

5. Practical applications

In order to study the performance of the various goodness-of-fit tests, the criteria of § 4 were tried on the observed data of four triclinic crystals (Table 1). From the $|F_o|$ data the reflexions for which $\sin\theta/\lambda \le 0{\cdot}55$ were partioned into cells of $\sin\theta/\lambda$ (abbreviated as s). Each unobserved reflexion was also included* with a value which is half the local least observed value of the structure amplitude. The values† of $\langle I_o\rangle$ and that of s for each cell were then obtained. Least-squares polynomials of different degrees of the type

$$\langle I_o\rangle = \sum_{i=0}^{N} a_i s^i \tag{20}$$

were fitted to this set of points $(s_i, \langle I_o\rangle_i)$, $i=1$ to m. The polynomial equation which gives the best fit with the data was then determined from a study of the standard deviations for the various degrees N of the polynomial equation. The coefficients of the best polynomial and its degree are given in Table 2 (for the principle of the method and the programming see Carnahan, Luther & Wilkes, 1969). The graphs of the best polynomial equations for the four crystals are shown in Fig. 1.

From the reflexions of set A, reflexions for which $s < 1/a_{\min}$ (where $a_{\min}$ is the smallest cell dimension of the crystal) were omitted (Wilson, 1949). The z values (note that $z = I_o/\langle I_o\rangle$) of all the reflexions (including the unobserved ones) in the range $(1/a_{\min}) \le \sin\theta/\lambda \le 0{\cdot}5$ (for convenience, this set of reflexions is called set B) were then obtained from the analytical expression for $\langle I_o\rangle$ (see equation 20). The value of $\langle z\rangle$ for the reflexions of set B was then calculated and is close to unity as expected* (Table 1).

The binomial test was conducted with the reflexions of set B† by taking the null hypothesis to be that the crystal is C. The results obtained for the various crystals are summarized in Table 3. Here the values in the row OV(C) represent the observed values of the statistic when H_0 is that the crystal is C and those in the row CV(C) are the critical values of the statistic corresponding to $\alpha = 0{\cdot}05$ when H_0 is that the crystal is C. A comparison of the corresponding values in these rows shows that this test leads to a correct decision in all the cases at the level $\alpha = 0{\cdot}05$.

The other five tests were conducted with the data with a y cut-off. From a study of the least observed values of $|F_o|$ in the various regions of $\sin\theta/\lambda$ for the four crystals, it was found that a value of 0·15 for y_t would be a suitable cut-off limit. From the set B, the reflexions for which $y < 0{\cdot}15$ were therefore omitted, resulting in a new set C. With the reflexions of set C, the five tests‡ were conducted and the results obtained are summarized in Table 3.

* Such a procedure is essential for deriving z values as accurately as possible. However, the unobserved reflexions could be subsequently cut off in the actual tests by using the modified pdf of y (or z). Note also that for convenience we shall call the set of reflexions (including the unobserved ones) for which $0 \le s \le 0{\cdot}55$ set A.

† If the extreme values of $\sin\theta/\lambda$ corresponding to the reflexions within a thin shell in reciprocal space are s_1 and s_2, then the value of s corresponding to this shell is the weighted average given by $(\frac{3}{4})(s_2^4 - s_1^4)/(s_2^3 - s_1^3)$. This result can be derived analytically by subdividing the given shell into elementary subshells (Parthasarathi, 1975).

* Though the relation $\langle z\rangle = 1{\cdot}0$ does not provide a test for centrosymmetry, it provides a check on the accuracy of the z data. Hence the evaluation of $\langle z\rangle$ is recommended in statistical tests.

† It is obvious that the inaccuracy of the z values of the unobserved reflexions will not affect this test since all the reflexions are grouped into only two classes, namely, M and $\overline{M}$ [see § 4 (iii)].

‡ Reflections of set C evidently satisfy the conditions that $(1/a_{\min}) \le (\sin\theta/\lambda) \le 0{\cdot}5$ and $y_t \le y \le \infty$ simultaneously. It is obvious that this set would contain no unobserved reflexions.

Table 2. *Coefficients a_i of the best polynomial equation $\langle I_o\rangle = \sum_{i=0}^{N} a_i s^i$ for the various crystals*

Crystal	I	II	III	IV
N	5	4	6	6
a_0	0·6159700 (3)*	0·8896042 (3)	0·7032594 (3)	0·4816834 (3)
a_1	0·5251690 (4)	0·7415506 (4)	0·9569445 (4)	0·8114339 (4)
a_2	−0·1383913 (5)	−0·7500976 (5)	−0·9625175 (5)	−0·8994018 (5)
a_3	0·7345037 (4)	0·1926998 (6)	0·2123507 (6)	0·2701496 (6)
a_4	0·8099037 (5)	−0·1581581 (6)	0·1661707 (6)	−0·1858583 (6)
a_5	−0·1085313 (6)	–	−0·1028943 (7)	−0·3570500 (6)
a_6	–	–	0·8855826 (6)	0·4455345 (6)

* 0·6159700 (3) means $0{\cdot}6159700 \times 10^3$ *etc.*

From Table 3 we obtain the following results when H_0 is that the crystal is C: when the level of significance $\alpha=0{\cdot}05$, the chi-square test, the likelihood-ratio test, the binomial test, the Kolmogorov test and Neyman–Barton test (*via* statistic p_2^2) lead to correct decisions, while the Smirnov–Cramer–Von Mises test fails for II and the Neyman–Barton test (*via* statistic p_1^2) fails for II and IV.

In order to make the study complete the various tests were repeated by taking the null hypothesis to be that the crystal is NC. The observed values of the statistic thus obtained are summarized in Table 3 against the rows marked OV(NC). The critical values (when $\alpha=0{\cdot}05$) for all the tests except the binomial test turn out to be the corresponding ones obtained for the C case and hence these correspond to the values given against the rows marked CV. The critical values (when $\alpha=0{\cdot}05$) for the binomial test for the present case are given against the row CV(NC). From Table 3, we obtain the following results when H_0 is that the crystal is NC: (i) the chi-square test and the likelihood-ratio test lead to correct decisions in all cases ($\alpha=0{\cdot}05$); (ii) the binomial test fails in IV and the Neyman–Barton smooth test (*via* statistic p_2^2) in III (when $\alpha=0{\cdot}05$); (iii) each of the Kolmogorov test, the Smirnov–Cramer–Von Mises test and the Neyman–Barton

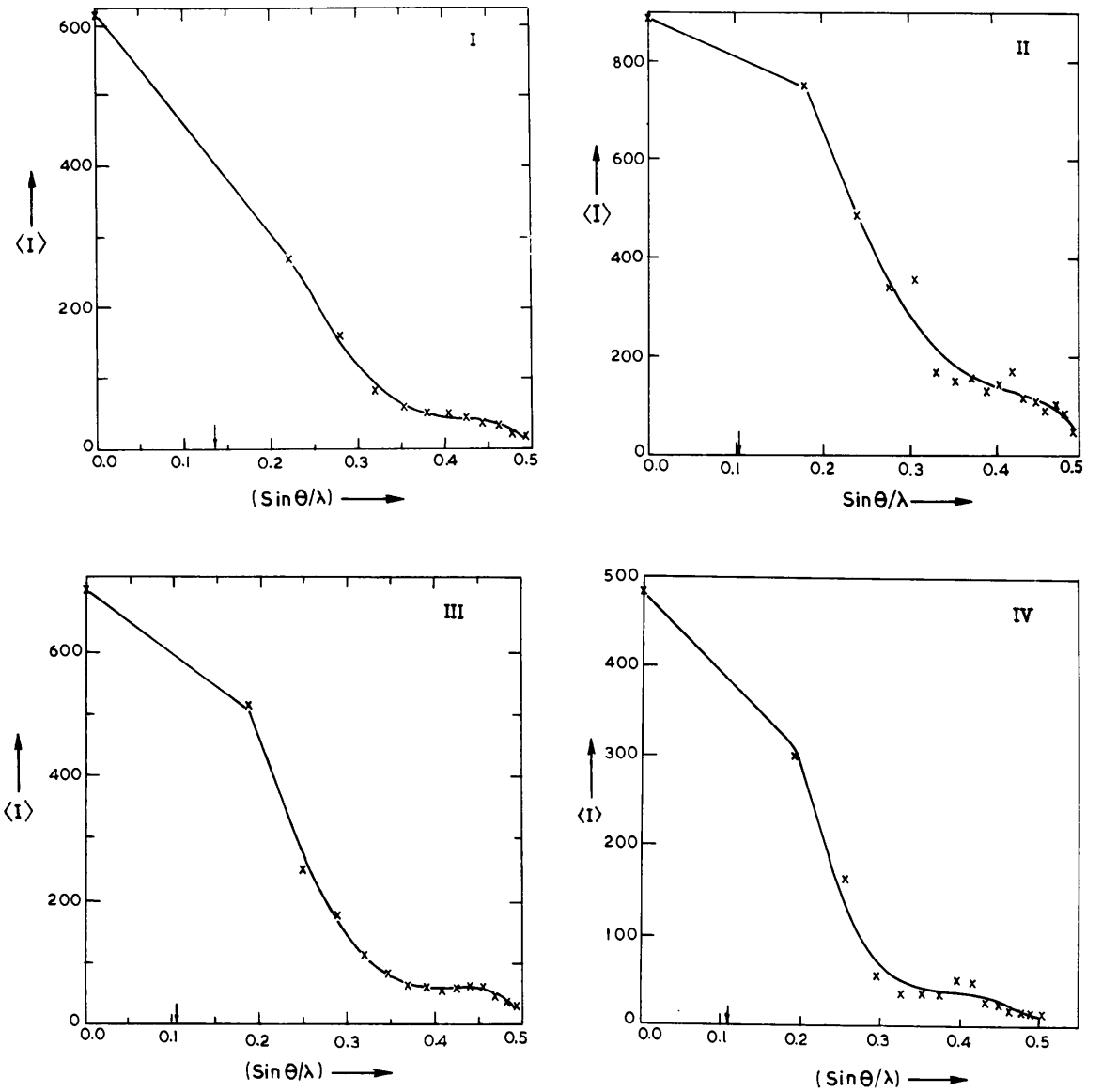

Fig. 1. The best least-squares polynomial equation for $\langle I_o \rangle$ as a function of s for the four crystals (I) to (IV) of Table 1 (solid line). The crosses correspond to the observed data. The arrow denotes the least value of s for the reflexions of set B, namely $1/a_{\min}$ (see § 5 of the text for definition of set B).

smooth test (*via* statistic p_1^2) fails for two crystals.

From the above discussion and a study of Table 3 it is seen that when the two observed values of any statistic (namely those corresponding to OV(C) and OV(NC) are compared with the corresponding theoretical critical values (at $\alpha=0{\cdot}05$ level) expected for the C and NC cases (namely, those corresponding to CV(C) and CV(NC) in the binomial test and that corresponding to CV in the other tests) a correct decision could always be reached in all tests (except the Neyman–Barton smooth test *via* statistic p_1^2) by rejecting that alternative for which the observed value of the statistic lies further into the critical region. If, for example, the Neyman–Barton smooth test (*via* statistic p_2^2) is taken for III, it is seen that while the value of 74·35 for p_2^2 lies far in the critical region (*viz.* $5{\cdot}99 \leq p_2^2 < \infty$) the value of 6·07 for p_2^2 lies at the border of the critical region. It is thus obvious that, although this test just fails at $\alpha=0{\cdot}05$ level when H_0 is that the crystal is NC, it follows that a correct decision (*viz.* the crystal is NC) could be obtained by the above procedure by rejecting the worse offender. Thus from the above study it turns out that a safe procedure for conducting the tests would be to compute the two observed values of the statistic in each case and compare them with the corresponding critical values (for $\alpha=0{\cdot}05$, say) expected for the C and NC cases and finally reject the one for which the fit is worse.

6. Concluding remarks

Though the various criteria were discussed in §§ 4 and 5 by assuming the crystal to obey the requirements of the basic Wilson (1949) distributions these could also be used to test the goodness-of-fit for crystals containing a few heavy atoms (besides a large number of light atoms) and for crystals exhibiting pseudosymmetry in the atomic arrangement, provided the appropriate distribution functions are available. For triclinic crystals containing one or two heavy atoms in the unit cell the distributions and other relevant results are available (Sim, 1958; Srinivasan, 1960; Srikrishnan & Parthasarathy, 1970). The distribution functions needed for testing the goodness-of-fit for crystals exhibiting pseudosymmetry are also available for a few specific cases (Lipson & Woolfson, 1952; Rogers & Wilson, 1953). The procedure for conducting the tests is similar to that discussed in § 5. For structures with heavy atoms the pdf of y depends on the parameter σ_1^2 (the fractional heavy-atom contribution to the local mean intensity) which is a function of s and hence it is essential to use the mean value of σ_1^2 for specifying the distribution function of y.

One of the authors (MNP) thanks the University Grants Commission, New Delhi, India for financial assistance.

Table 3. *Results of the various statistical tests on the crystals of Table* 1

The case where a test leads to a wrong result (at $\alpha=0{\cdot}05$ level) is indicated by showing the corresponding observed value of the statistic in bold type.

Test	Test statistic*		Crystal† I	II	III	IV
1. Chi square	T	k‡	51	57	60	59
		CV	67·22	74·18	77·65	76·50
		OV(C)	52·87	73·16	126·37	149·73
		OV(NC)	182·05	229·61	60·99	66·94
2. Likelihood ratio	Λ	k	51	57	60	59
		CV	67·22	74·18	77·65	76·50
		OV (C)	52·22	71·66	137·91	155·07
		OV (NC)	149·83	182·95	59·95	71·21
3. Binomial	B	CV (C)	1·65	1·65	1·65	1·65
		OV (C)	−1·35	−4·45	11·78	11·23
		CV (NC)	−1·65	−1·65	−1·65	−1·65
		OV (NC)	−15·02	−21·39	−1·26	**−1·80**
4. Kolmogorov	$n^{1/2}D_n$	CV	1·36	1·36	1·36	1·36
		OV (C)	0·73	1·21	2·73	2·80
		OV (NC)	2·37	2·18	**1·65**	**1·38**
5. Smirnov–Cramer	nW^2	CV	0·46	0·46	0·46	0·46
		OV (C)	0·08	**0·53**	2·10	2·28
		OV (NC)	2·12	2·31	**0·51**	**0·51**
6. Neymon–Barton smooth	(*a*) p_1^2	CV	3·84	3·84	3·84	3·84
		OV (C)	0·55	**5·71**	**4·37**	**3·80**
		OV (NC)	5·30	**0·91**	3·56	**4·11**
	(*b*) p_2^2	CV	5·99	5·99	5·99	5·99
		OV (C)	0·57	5·73	74·35	80·18
		OV(NC)	80·94	104·24	**6·07**	4·83

* For definition of these see § 4.

† I and II are centrosymmetric and III and IV are non-centrosymmetric.

‡ k=number of equiprobable histogram bins for the y data of set C.

References

BIANCHI, R., MOROSI, G., MUGNOLI, A. & SIMONETTA, M. (1973). *Acta Cryst.* B**29**, 1196–1208.
CARNAHAN, B., LUTHER, H. A. & WILKES, J. O. (1969). *Applied Numerical Methods*. New York: John Wiley.
COCHRAN, W. G. (1952). *Ann. Math. Statistic.* **23**, 315–345.
COCHRAN, W. G. (1954). *Biometrics*, **10**, 417–451.
EADIE, W., DRIJARD, D., JAMES, F. E., ROOS, M. & SADOULET, B. (1971). *Statistical Methods in Experimental Physics*. Amsterdam: North Holland.
HARGREAVES, A. & GOGOI, B. N. (1966). *Acta Cryst.* **21**, 26–28.
HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210–214.
KARLE, I. L. (1969). *Acta Cryst.* B**25**, 2119–2126.
KISTENMACHER, T. J., HUNT, D. J. & MARSH, R. E. (1972). *Acta Cryst.* B**28**, 3352–3361.
KROON, J. & KANTERS, J. A. (1973). *Acta Cryst.* B**29**, 1278–1283.
LIPSON, H. & WOOLFSON, M. M. (1952). *Acta Cryst.* **5**, 680–682.
PARTHASARATHI, V. (1975). Ph. D. Thesis, Univ. of Madras.
RAMACHANDRAN, G. N. & SRINIVASAN, R. (1959). *Acta Cryst.* **12**, 410–411.
ROGERS, D., STANLEY, E. & WILSON, A. J. C. (1955). *Acta Cryst.* **8**, 383–393.
ROGERS, D. & WILSON, A. J. C. (1953). *Acta Cryst.* **6**, 439–449.
SIEGEL, S. (1956). *Non-parametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
SIM, G. A. (1958). *Acta Cryst.* **11**, 123–124.
SRIKRISHNAN, T. & PARTHASARATHY, S. (1970). *Z. Kristallogr.* **131**, 186–195.
SRINIVASAN, R. (1960). *Acta Cryst.* **13**, 388–394.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1977). A**33**, 844–846

The Calculation of Electron Diffraction Intensities by the Multislice Method

BY G. R. ANSTIS

Electron Microscope Unit, The University of Sydney, NSW, 2006, *Australia*

(*Received* 30 *November* 1976; *accepted* 26 *April* 1977)

The calculation of wave functions of scattered electrons by the multislice method of Cowley and Moodie with a finite number of beams is shown to lead to the solution of a finite, closed set of differential equations in the limit that the slice thickness approaches zero. The solution is normalized but differs from the exact wave function unless sufficient beams are included in the calculation. Hence, normalization is not sufficient to ensure that the computed wave function equals the exact wave function. The implications of this result for numerical work are discussed.

Introduction

The multislice method of Cowley and Moodie has been shown by Goodman & Moodie (1974) to give the solution of the form of Schrödinger's equation in which backscattering is neglected. If one imposes the condition that the electrons are scattered by a periodic potential then this equation is equivalent to a countably infinite set of coupled differential equations describing the amplitudes and phases of diffracted beams as a function of position in the scatterer.

Since numerical methods of solving Schrödinger's equation can account for the effects of only a finite number of beams it is worth examining the properties of approximate solutions obtained by methods involving a finite number of beams, and to see how the accuracy of these solutions might be estimated. While it is often not possible to determine analytic solutions to problems in which the effects of three or more beams are important, certain properties of the solutions can be obtained.

One method of determining approximate solutions to Schrödinger's equation is to consider a closed, finite subset of the differential equations which are equivalent to Schrödinger's equation. It can then be shown that the solution of this finite set of equations is normalized, a term which is defined later and which implies that the number of electrons incident on the scatterer equals the number leaving it. This is also a property of the exact solution of Schrödinger's equation. To determine whether the solution is an adequate approximation to an unknown exact solution it must be compared with the solution to a different set of differential equations.

Goodman & Moodie (1974) have suggested that the multislice method has an advantage over the method of truncating the set of differential equations in that it is possible to determine the accuracy of a wave function obtained by the multislice method with a finite number of beams without having to compare the results of calculations with different numbers of beams. Since for any non-zero slice thickness it is only in the limit that the number of beams becomes infinitely large that a multislice calculation results in