Acta Cryst. (1971). A**27**, 536

Dispersion Analysis of Single-Crystal Diffractometer Measurements. I. Methods of Investigation

By H. Schulz

Institut für Kristallographie und Petrographie of the Eidgenössische Technische Hochschule, Sonneggstrasse 5, 8006 *Zürich, Switzerland*

and P. J. Huber

Lehrstuhl für Wahrscheinlichkeitsrechnung und mathematische Statistik, Eidgenössische Technische Hochschule, Clausiusstrasse 1/3, 8006 *Zürich, Switzerland*

(*Received* 27 *July* 1970 *and in revised form* 23 *November* 1970)

It is shown that the distribution of symmetry-equivalent reflexions around their mean value can be compared with their expected distribution using the χ^2 test. Three methods are offered to investigate any part of a data set for significant differences in the variance of the counting statistics and thereby for systematic errors. These methods are applied to a test example. Also, the distribution of repeated measurements of pulse rates is investigated.

1. Introduction

In diffraction experiments with single crystals the intensity I of a diffracted beam can be measured by counting the pulses. The single measurements (R) are distributed around their expected value $E(R)$ according to the Poisson statistics. For larger counting rates (>100 counts) the Poisson distribution becomes a special normal distribution in which the variance $V(R)$ is equal to the expected value $E(R)$. Such counting rates are normally used for intensity measurements with single-crystal diffractometers. The intensity I of a reflexion is calculated from normally distributed pulse rates (R_n) using the linear combination

$$I=\sum_n^m a_nR_n\,. \qquad (1)$$

Therefore, I is normally distributed too. The variance $V(I)$ can also be calculated from the variances $V(R_n)$ (Linder, 1964):

$$V(I)=\sum_n^m a_n^2V(R_n)\simeq\sum_n^m a_n^2R_n\,. \qquad (2)$$

The intensity measurements of symmetry-equivalent reflexions can be interpreted as a pseudo-repetition of the intensity measurement of one single reflexion. The mean value $\bar{I}$ and the estimated variance $V'(I)$ (square of the estimated standard deviation) of a single measurement I can be calculated from a group of symmetry-equivalent reflexions (K members) using equations (3) and (4):

$$\bar{I}=\sum_n^k \frac{I_n}{k} \qquad (3)$$

$$V'(I)=\frac{1}{k-1}\sum_n^k(\bar{I}-I_n)^2\,. \qquad (4)$$

The expected value of the quotient

$$\frac{V'(I)}{\frac{1}{k}\sum_n^k V(I_n)}\simeq\frac{V'(I)}{\frac{1}{k}\sum_n^k a_n^2R_n}=\frac{V'(I)}{V''(I)}=Y \qquad (5)$$

is approximately equal to 1, if the deviations from $E(I)$ are due to the counting statistics only. $V''(I)$ is designated as the expected variance of a single measurement I in contrast to the estimated variance $V'(I)$ of equation (4). Y may vary over some orders of magnitude (Abrahams, 1964), because k, the number of measurements, is normally too low for the purposes of statistics. Therefore, it is impossible to use Y for the analysis of systematic errors. To avoid these difficulties the χ^2 test may be applied. As shown in §§ 2 & 3 one χ^2 value corresponds to each group of symmetry-equivalent reflexions and any number of these groups can be put together for only one statement.

2. χ^2 test of repeated intensity measurements of single-crystal reflexions

The intensity of a single-crystal reflexion may be measured k times with a counting tube. Peak count rates (P_n) and background count rates (B_n) have equal measuring time. P_n and B_n are Poisson-distributed. (In the notation of equation (1) this means $R_1=P_n$, $R_2=B_n$, $a_1=1$, $a_2=-1$.) Furthermore it is valid that

$$E(P_n)=\lambda\,, \qquad (6)$$

$$E(B_n)=\mu\,, \qquad (7)$$

$$I_n = P_n - B_n \tag{8}$$

and

$$W_n = P_n + B_n \,. \tag{9}$$

The expected values and the variances of I_n and W_n follow from equations (6) to (9):

$$E(I_n) = \lambda - \mu \,, \tag{10}$$

$$E(W_n) = V(I_n) = V(W_n) = \lambda + \mu \,. \tag{11}$$

$\bar{I}$ and Y are calculated from I_n according to equations (3) and (5):

$$Y = \frac{\frac{1}{k-1}\sum_{n=1}^{k}(I_n - \bar{I})^2}{\frac{1}{k}\sum_{n=1}^{k} W_n} = \frac{V'(I)}{V''(I)} \,. \tag{12}$$

The expected value of $V'(I)$ is given by equation (11), and the variance

$$V[V'(I)] = (\lambda + \mu)^2 \frac{2}{k-1} \,. \tag{13}$$

$V''(I)$ has the expected value (11) and from equation (11) the variance is:

$$V[V''(I)] = \frac{\lambda + \mu}{k} \,. \tag{14}$$

$V'(I)/(\lambda + \mu)$ and $V''(I)/(\lambda + \mu)$ have the expected value 1 [*cf.* equation (11)]. The corresponding variances are:

$$V\left(\frac{V'(I)}{\lambda + \mu}\right) = \frac{k-1}{2} \,, \tag{15}$$

$$V\left(\frac{V''(I)}{\lambda + \mu}\right) = \frac{1}{(\lambda + \mu)k} \,. \tag{16}$$

The right side of equation (16) is nearly zero for large values $(\lambda + \mu)$. Such values are obtained from intensity measurements of single-crystal reflexions. Therefore, Y has an approximate χ^2 distribution with $(k-1)$ degrees of freedom.

3. Application of the χ^2 test

About a thousand intensity values are usually measured from a single crystal with a diffractometer. These measurements can be grouped in symmetry-equivalent reflexions (S groups). The k $(k \geq 2)$ intensity measurements I_n of an S group may be understood as a k-times repeated measurement of one reflexion intensity value. With equation (12) a $\chi^2(S)$ with $(k-1)$ degrees of freedom can be calculated for each S group. These $\chi^2(S)$ values correspond to upper significance levels (USL),* which can be read from χ^2 tables. It is not possible to compare the agreement between the expected $V''(I)$ and the estimated variances $V'(I)$ with only one $\chi^2(S)$. S groups may be ordered to property groups (P groups) according to certain view (*i.e.* increasing intensities). A $\chi^2(P)$ belongs to each P group. The weight of a statement, drawn from a $\chi^2(P)$, increases with an increasing number of measurements of a P group. S groups with only two reflexions can be used, because the large dispersion of $V'(I)$ (*cf.* equation (4); Jeffery, 1964) has no effect due to the summarizing (*cf.* equation 17). The following methods may be used for the statistical analysis of a P group:

(*a*) The $\chi^2(S)$ values of all S groups (p members) are added and thus lead to the $\chi^2(P)$ value:

$$\chi^2(P) = \sum_{n}^{p} \chi_n^2(S) \,. \tag{17}$$

The degree of freedom of $\chi^2(P)$ equals the sum of the degrees of freedom of the p P-groups:

$$G = \sum_{n}^{p} (k_n - 1) \,. \tag{18}$$

An USL for the whole P group can be determined from this $\chi^2(P)$ value. An approximation formula may be used (Abrahams, 1969) for G's not included in the tables.

(*b*) If the intensities of the S groups are normally distributed, $\chi^2(P)$ belongs to a normal distribution with the mean G and the variance $2G$ (van der Waerden, 1965). The hypothesis of the agreement of the expected variances $V''(I)$ [*cf.* equation (5)] with the estimated variances $V'(I)$ [*cf.* equation (4)] can be checked by comparing $|\chi^2(P) - G|$ and $\sigma[\chi^2(P)] = \sqrt{2G}$. The hypothesis should be rejected, if the differences are larger than $2\sigma[\chi^2(P)]$ or $3\sigma[\chi^2(P)]$. The probability of agreement is about 0·06 in the first case and about 0·01 in the second case.

(*c*) The USL's of the S groups are distributed uniformly over the range $0-1$ whenever the expected and estimated variances agree. The USL's are accumulated near zero, if the estimated variances are larger than the expected ones. Each $\chi^2(P)$ falls into a fixed interval corresponding to its USL, if the range $0-1$ is divided into equal parts. Consider the interval divided into 10 parts; then the expected value for each part is equal to $p/10$. The hypothesis of the uniform distribution of the $\chi^2(P)$ should be rechecked by a χ^2 test. The degree of freedom of this test equals the number of intervals minus 1, because the number of measured and expected events must agree; thus, one linear relation exists.

Method (*a*) contains the full information of the χ^2 test. Method (*b*) allows an estimation that is sufficient for most of the practical applications and can be easily programmed for a computer, because values of the USL are not needed. Method (*c*) is not influenced by extremely large or small χ^2 values, but only by the number of such extreme values.

4. Test example

The intensities were measured with an off-line automatic single-crystal diffractometer of the Siemens Company using Nb-filtered Mo radiation.

* Upper significance level: possibility of the upper tail of the χ^2 distribution.

A reflexion intensity of a single crystal was measured 372 times during 8 hours immediately after starting the instrument. The peak intensity was measured with the θ–2θ scan method, while the background was measured in a fixed position right and left of the peak with equal measuring time for peak and background. As Fig. 1 shows the expected and the measured distribution do not agree.

The hypothesis of agreement is clearly rejected ($\chi^2 = 268$, 10 degrees of freedom). It is a reasonable assumption that the result is due to a long time drift. To check this hypothesis five measurements following each other were collected in one group. In this way an S group is simulated. Each group of 5 intensities, 74 members in all, belongs to one P group.

A $\chi^2(S)$ value was calculated and the corresponding USL was read from a table for each group of five intensities. Then the hypothesis of a long time drift was checked with the three methods described in § 3. $\chi^2(P)$ and G equal 338 and 296, respectively. An USL of 0·08 corresponds to these values. The hypothesis of a long time drift may be accepted. Standard deviation $\sigma[\chi^2(P)]$ equals $\sqrt{2 \cdot 296} = 24{\cdot}2$. The difference $|\chi^2(P) - G| = 42$ is less than $2\sigma[\chi^2(P)]$. This result allows the checked hypothesis to be accepted. The results of the χ^2 test of the χ^2 distribution are listed in Table 1. The corresponding USL of 0·60 is much higher than expected from the results of the other two methods. It is caused by the suppression of some extremely large χ^2 values.

Table 1. *χ^2 test of the χ^2 distribution for the groups of 5 intensities*

Upper significance level	Experimental values	Expected values	Difference	Quotient
0·0 < 0·1	10	7·4	2·6	6·8/7·4 = 0·90
0·1 < 0·2	6	7·4	−1·4	2·7/7·4 = 0·36
0·2 < 0·3	9	7·4	1·6	2·5/7·4 = 0·34
0·3 < 0·4	8	7·4	0·6	0·4/7·4 = 0·05
0·4 < 0·5	8	7·4	0·6	0·4/7·4 = 0·05
0·5 < 0·6	9	7·4	1·6	2·5/7·4 = 0·34
0·6 < 0·7	2	7·4	−5·4	29·0/7·4 = 3·92
0·7 < 0·8	9	7·4	1·6	2·5/7·4 = 0·34
0·8 < 0·9	6	7·4	−1·4	2·7/7·4 = 0·36
0·9 < 1·0	7	7·4	−0·4	0·2/7·4 = 0·02
Totals	74	74		$\chi^2 = 6{\cdot}68$

Degree of freedom = 10 − 1 = 9.
Upper significance level = 0·60

5. Inspection of the distribution function of the utilized single-crystal diffractometer

Pulse rates should be distributed according to the Poisson statistics. For larger pulse rates the Poisson distribution changes into a special normal distribution in which the mean is equal to the variance. Huber (1968) pointed out that so called natural normal distributions do not follow a Gauss curve exactly, because events with low probability are observable too often. This effect may be roughly compensated for by an increase in the expected variance.

The distribution function of the utilized diffractometer was investigated by repeated measurements of different pulse rates. The instrument was turned on three days before starting the measurements. To reduce the influence of power-supply fluctuations, pulse rates were measured between 1 and 4 a.m.

The mean of the lowest pulse rate was 4·9 pulses/measurement. Fig. 2 shows the measured and the expected distribution. A χ^2 test of the distribution (Linder, 1964) results in an USL of 0·7; the hypothesis of agreement may be accepted.

The larger pulse rates were determined by intensity measurements of a single-crystal reflexion as described in § 4. The distribution of the intensities around their mean value is shown in Fig. 3. A χ^2 test was applied to this distribution (Table 2). Fig. 3 shows that more events

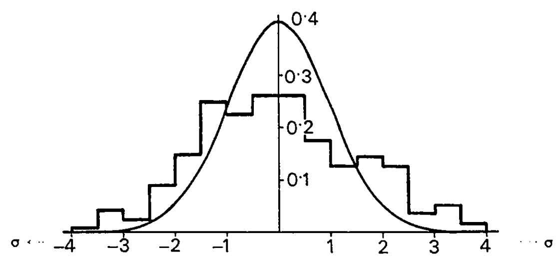

Fig. 1. Measured and expected distribution of the intensities of a single-crystal reflexion drawn as standard normal distribution. Unstable state of the diffractometer. (Mean values from 372 measurements: peak 41·134, background 5·321, intensity 27·254 counts.)

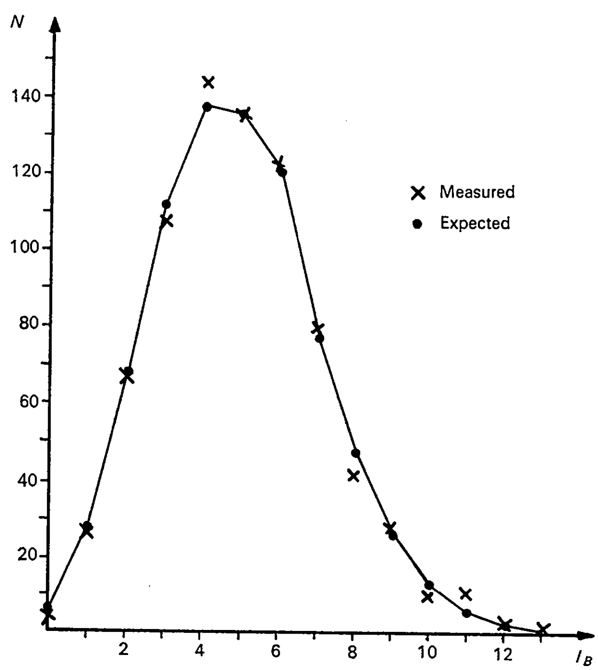

Fig. 2. Measured and expected distribution of 783 pulse rates with a mean value of 4·9 pulses/measurement.

Table 2. χ^2 *test of the intensity distribution*

Intensity measurement with the diffractometer in stable state. $N=179$ measurements; intensity mean value, $\bar{I}=27{\cdot}249$; estimated standard deviation $\sigma(I)=195$.

Intensity classes $\bar{I}+n\sigma<I\leqq\bar{I}+(n+1)\sigma$	N(measured)	N(expected)	Difference	Quotient
$n=-3$	9	4	+5	25/4=6·25
−2	21	24	−3	9/24=0·37
−1	58	61	−3	9/61=0·15
0	62	61	+1	1/61=0·02
1	22	24	−2	4/24=0·17
2	4	4	+2	4/4 =1·00
3	2	0		
	Totals 178	178		$\chi^2=7{\cdot}96$

Degree of freedom = 6 − 2 = 4.
Upper significance level = 0·09.

fall in the outer classes than expected from the normal distribution. The χ^2 test ($\chi^2=7{\cdot}96$, 4 degrees of freedom, USL = 0·09) revealed that the variance calculated from the mean value of the intensities may be too small. These results are strongly influenced by the few measurements of the outer classes. It was then decided to calculate χ^2 values from 179 peak, background and intensity measurements using equation (12). These values were 241, 236, and 221 with a degree of freedom $G=179-1=178$. In all three cases $|\chi^2-G|$ is larger than $2\sigma(\chi^2)=2\sqrt{356}=36{\cdot}8$. This and further investigations (*cf.* § 3 of part II of this paper) show that it is necessary to increase the expected variance, *i.e.* the variance calculated from the pulse rates [*cf.* equations (2) and (5)].

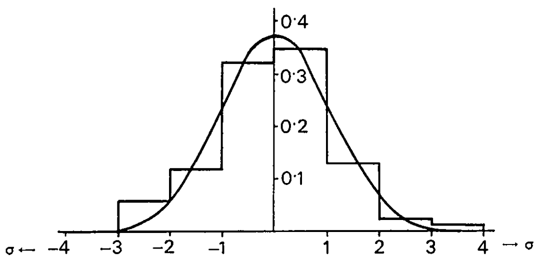

Fig. 3. Measured and expected distribution of the intensities of a single-crystal reflexion drawn as standard normal distribution. Stable state of the diffractometer. (Mean values from 179 measurements: peak 32·564, background 5·310, intensity 27·254 counts.)

References

ABRAHAMS, S. C. (1964). *Acta Cryst.* **17**, 1327.
ABRAHAMS, S. C. (1969). *Acta Cryst.* A**25**, 165.
HUBER, P. J. (1968). *Mathematical Centre Tracts, Amsterdam*, **27**, 1.
JEFFERY, J. W. (1964). *Acta Cryst.* **17**, 1329.
LINDER, A. (1964). *Statistische Methoden*, pp. 359, 368. Basel und Stuttgart: Birkhäuser Verlag.
WAERDEN, B. L. VAN DER (1965). *Mathematische Statistik*, p. 100. Berlin: Springer Verlag.