

gether into one infinite layer through the oxygen atoms parallel to (100). In addition, each oxygen atom belongs to one Hg_I octahedron, and each chlorine atom to two such octahedra. It follows that the structure is built up of alternating layers of Hg_I^{++} cations and polymeric $(\text{OHgCl})^-$ anions. These layers can be best realized from the projection on (010) plane in (Fig. 6). The Hg_{II} tetrahedron is far more distorted than the Hg_I octahedron, as may be seen from Table 3. This distortion is partly caused by the polarization action of the mercuric ion (Hg_{II}) and partly by the packing conditions dictated by the largest Cl ions. The O-O approaches in the same tetrahedron are 4.14, 3.90 and 2.48 Å. The last value belongs to the common edge of two linked tetrahedra, since each tetrahedron shares one of its O-O edges with its neighbour. That is the shortest distance between the non-bonded atoms.

The author wishes to express his sincere gratitude to Dr D. Grdenić (Laboratory for General and Inorganic Chemistry, Faculty of Science) for many valuable suggestions and helpful discussions and encouragement during the whole work. Thanks are due to Prof. M. Pačić at whose laboratory (Physical Institute, Fac-

ulty of Science) a part of the experimental work was performed. The investigation described in this paper was supported by the Institute Rudjer Bošković, Zagreb.

References

- ARCTOWSKI, H. (1895). *Z. anorg. Chem.* **9**, 178.
 AURIVILLIUS, K. (1954). *Acta Chem. Scand.* **8**, 523.
 BLASS, J. (1879). *Tscherm. Mitt.* **2**, 177.
 CAROZZI, E. (1926). *Gazz. chim. ital.* **56**, 175.
 GARRET, A. B. & LEMLEY, J. (1942). *J. Amer. Chem. Soc.* **64**, 2380.
 GAWRICH, S. (1938). *Roczn. Chem. (Ann. Soc. Chim. Polonorum)*, **18**, 107. (Quoted in *Chem. Zbl.* (1940), **1**, 3622.)
 GRDENIĆ, D. (1952a). *Acta Cryst.* **5**, 367.
 GRDENIĆ, D. (1952b). *Acta Cryst.* **5**, 238.
 GRDENIĆ, D. & ŠČAVNIČAR, S. (1953). *Nature, Lond.* **172**, 584.
Internationale Tabellen zur Bestimmung von Kristallstrukturen (1935). Berlin: Borntraeger.
 NEST, J. S. v. (1910). *Z. Kristallogr.* **47**, 263.
 PAULING, L. (1940). *The Nature of the Chemical Bond*. Ithaca: Cornell University Press.
 SCHOCH, E. P. (1903). *Amer. Chem. J.* **29**, 319. (Quoted in *Chem. Zbl.* (1903), **1**, 1173.)
 TARUGI, N. (1901). *Gazz. chim. ital.* **31**, 313. (Quoted in *Chem. Zbl.* (1901), **2**, 1147.)

Acta Cryst. (1955). **8**, 383

The Probability Distribution of Intensities. VI. The Influence of Intensity Errors on the Statistical Tests

BY D. ROGERS, E. STANLEY* AND A. J. C. WILSON

Viriamu Jones Laboratory, University College, Cardiff, Wales

(Received 2 February 1955)

The statistical tests, three for centrosymmetry and one for planes of symmetry, have been examined to determine quantitatively their sensitivity to errors in the original intensity data. The errors considered are (i) random errors proportional to I , (ii) errors systematic in I of the form $I_0 = k\Sigma(1 - \exp(-I/k\Sigma))$, (iii) errors systematic in $\sin \theta$, and (iv) errors associated with the non-observance of very weak reflexions. Errors (i) and (ii) of ordinary magnitude produce no confusion. Errors (iii) can affect the tests only through errors in the determination of $\langle I \rangle$ as a function of $\sin^2 \theta$, and to these the tests are rather sensitive. Errors (iv) can also have marked effects, for which, however, allowances can be made. The tests are unequally affected by a given amount of error, but the results derived here permit the estimation in a given problem of likely outer limits for the result of each test. Results lying outside this limit may be regarded as significantly different and indicative of some structural peculiarity.

1. Introduction

1.1. Several authors have reported successful decisions for or against centrosymmetry using the statistical tests discussed in the preceding papers of this series, but results are sometimes obtained which do not agree well with the criteria which characterize the two ideal

intensity distributions, the centric and the acentric. The discrepancies may be attributed to:

(i) some violation of the assumptions underlying the derivation of the ideal distribution functions, e.g. too few atoms (for allowances see Wilson (1951), Hauptman & Karle (1953) and Karle & Hauptman (1953)), lack of generality in their positions, pseudosymmetry (for a possible example see Bragg, Howells

* Now at the National Research Council, Ottawa, Canada.

& Perutz (1952)), additional non-crystallographic symmetry (considered by Lipson & Woolfson (1952) and discussed in more detail by Rogers & Wilson (1953)), or a few dominant atoms (examples of which were discussed by Howells, Phillips & Rogers (1950));

(ii) a total number of independent intensities too small to make the statistical treatment valid;

(iii) an erroneous evaluation of $\langle I \rangle$ as a function of $\sin^2 \theta$; or

(iv) errors in the observed intensity data.

The results obtained by these tests have always hitherto been interpreted on the assumption that the effects of the usual errors are small, but this assumption, which was given only a preliminary discussion by Howells *et al.* (1950), demands the closer study and proper justification that are supplied here. This is all the more necessary in view of recent attempts to correlate the statistical anomalies with structural peculiarities. A statistical result given by one of these tests can only be regarded as significantly different from the ideal figure if the discrepancy exceeds the maximum attributable to errors, and the estimation of this has necessitated a quantitative study.

Similar remarks apply to the average-multiple test discussed by Rogers (1950), and a study of the sensitivity to errors of this test has also been included in this paper.

1.2. The experimental estimation of intensities, whether by photography or counting techniques, is always accompanied by some measure of both random and systematic errors, and it will be convenient in what follows to treat them separately. The effects of random errors are discussed in § 2. The discussion of systematic errors is split into three parts: § 3 deals with errors systematic in I only, § 4 with errors systematic in $\sin \theta$ only, and § 5 with the errors due to unobserved reflexions, an error which is both systematic in I and a function of $\sin \theta$. The treatment of § 4 applies, therefore, equally to those of the experimental errors that are systematic in $\sin \theta$ (included in (iv) above) and those introduced in the course of the statistical work by the use of wrong values of $\langle I \rangle$, as in (iii) above. Errors that vary erratically throughout reciprocal space can be regarded as random. Absorption in an irregular crystal provides an example; on averaging the intensities over annuli or shells in reciprocal space, the equivalent of random errors will occur in each shell, though the change from shell to shell will be an error systematic in $\sin \theta$ (§ 4.1 below).

1.3. The treatment of the effects produced by these different kinds of error is facilitated by the introduction of the quantities

$$g_n = \langle I_o^n \rangle / \langle I^n \rangle, \quad (1)$$

where I is the true value of the intensity of a reflexion,

I_o is its observed value and n has the values $\frac{1}{2}$, 1 or 2. We shall assume throughout this paper that I conforms to one or other of the ideal distributions. For these $\langle I^n \rangle$ has the values in Table 1.

Table 1

	$\langle I^{\frac{1}{2}} \rangle$	$\langle I \rangle$	$\langle I^2 \rangle$
Acentric	$(\pi\Sigma/4)^{\frac{1}{2}}$	Σ	$2\Sigma^2$
Centric	$(2\Sigma/\pi)^{\frac{1}{2}}$	Σ	$3\Sigma^2$

By analogy with our definition of z we may write

$$z_o = I_o / \langle I_o \rangle, \quad (2)$$

so that

$$z_o = \frac{I_o}{g_1 \langle I \rangle} = \frac{I_o z}{I g_1}. \quad (3)$$

There will not normally, therefore, be any simple relation between $P_o(z)$, $N_o(z)$, the observed values,* and $P(z)$, $N(z)$, details of which were published earlier for the ideal distributions.

The value obtained for the test ratio when based on observed intensities is

$$\rho_o = \langle I_o^{\frac{1}{2}} \rangle^2 / \langle I_o \rangle, \quad (4)$$

which may be written as

$$\frac{\rho_o}{\rho} = \frac{(g_{\frac{1}{2}})^2}{g_1}. \quad (5)$$

Similarly, the observed value of the specific variance† is

$$v_o = \frac{\langle I_o^2 \rangle}{\langle I_o \rangle^2} - 1, \quad (6)$$

whence,

$$\frac{v_o + 1}{v + 1} = \frac{g_2}{g_1^2}. \quad (7)$$

2. Random errors

2.1. The random errors in visual estimation of the intensities are substantially larger than those in counting techniques, and it is not often that the precision of visual estimation will justify allowances for hydrogen atoms and the anisotropy of the thermal motion of individual atoms. Such a degree of refinement, however, is justifiable with intensity data obtained from counting techniques. At least a half of the

* We define $N(z)$ as $\int_0^z P(z) dz$ so that the corresponding quantity $\int_0^{z_o} P(z_o) dz_o$ could equally well be written as $N(z_o)$. We write it as $N_o(z)$ as there are occasions later in the paper when it is desirable to treat $N(z)$, $N_o(z)$ as distinct functions of a common variable.

† The specific variance quoted throughout this and earlier papers is the variance of the intensities. It should be distinguished from the specific variance of the F 's, which is related to the test ratio thus:

$$v_F = \rho^{-1} - 1.$$

residue of 4.6% found by Cochran (1951) for the reliability index of the [010] zone of adenine hydrochloride seems to be due to the random errors of observation (see Cochran, 1950). For the best visual estimates ultimate values of the reliability index range from 10 to 15%.

The exact way in which the error is related to the intensity is a matter of some doubt. It is fairly certain that the middle range of intensities, derived from estimates made on every film in multiple-film techniques, would be least subject to error, and that both the weak and the strong intensities, which can only be measured on a limited number of films, would be less accurate.

Booth (1947), in the course of an investigation of the effects produced by the errors of intensity estimation on the atomic coordinates derived from Fourier maps, examined two independent sets of observed intensities for dibenzyl (Robertson, 1935; Jeffrey, 1947) and concluded that the errors in $|F|$ could be taken as random and independent of the observed magnitude. He suggested that errors in $|F|$ proportional to $|F|$ (i.e. errors in I proportional to I) might be more acceptable, and such a distribution was first studied by Booth & Britten (1948). Phillips (1951) analysed the results for the two estimations on dibenzyl and also two of his own separate estimations of the intensities of the reflexions in the [010] zone of ephedrine hydrochloride. He found in both cases a relation of the form

$$\langle (\Delta I)^2 \rangle^{\frac{1}{2}} \propto I^{\frac{2}{3}} \quad (8)$$

with some evidence for a smaller exponent in the middle of the intensity range. The work was carried out on the intensities after allowing for the Lorentz and polarization factors, but it would be preferable to have used the intensities as measured on the films. It seems reasonable, then, to suppose that the errors in the observed intensities, due to the random errors of estimation, are, very roughly, proportional to the measured intensities and this was adopted as a mathematically convenient basis for the present study.

2.2. We shall assume, therefore, the relationship*

$$I_o = I(1 + \Delta), \quad (9)$$

in which Δ is normally distributed, with $\langle \Delta \rangle = 0$, $\langle \Delta^2 \rangle = \sigma^2$, $\langle \Delta^4 \rangle = 3\sigma^4$, etc. From this we deduce, since there is no correlation between I and Δ ,

$$g_1 = 1, \quad (10a)$$

$$g_{\frac{1}{2}} = \langle (1 + \Delta)^{\frac{1}{2}} \rangle, \quad (10b)$$

* This implies a finite probability of finding negative values of I_o although they are physically impossible. If, as in the experimental work of § 2.3, such values are reckoned zero a slight distortion is introduced into the distribution of errors. The results obtained in this section do not allow for this distortion but the error it introduces is quite negligible for $\sigma \lesssim \frac{1}{2}$.

$$g_2 = 1 + \sigma^2. \quad (10c)$$

Hence,

$$v_o = v + (1 + v)\sigma^2, \quad (11)$$

and

$$\varrho_o/\varrho = \langle (1 + \Delta)^{\frac{1}{2}} \rangle^2. \quad (12a)$$

This last equation can be expressed as the asymptotic series

$$\varrho_o/\varrho = \left\{ 1 - \frac{\sigma^2}{2^3} - \frac{15\sigma^4}{2^7} - \frac{315\sigma^6}{2^{10}} - \dots \right\}^2, \quad (12b)$$

which is suitable for numerical computation only when σ is small ($\lesssim 0.5$), and may then be written as

$$\varrho_o/\varrho = 1 - \sigma^2/4. \quad (12c)$$

The evaluation of $P_o(z)$ and $N_o(z)$ also is not straightforward. For example, we may write the combined probability of finding an observed intensity I_o , the result of all possible transitions Δ from all values of I , as

$$P(I_o)dI_o = \int P(I)P(\Delta)d\Delta dI. \quad (13)$$

These may be re-expressed in terms of I_o , I or I_o , Δ , but the integrals so obtained have not been evaluated in any convenient form. To overcome the difficulties in both this and the ratio test we have adopted an experimental approach to this problem which is discussed in the next paragraph.

2.3. An experimental investigation.

2.3.1. The introduction of random errors increases the dispersion of the intensities. The seriousness of this effect was investigated by introducing graded random errors into two sets of intensity data, one for each type of distribution, which otherwise appear to conform to the ideal distributions. They were:

(1) Centric distribution. Reflexions in the [010] zone of 1:2:3:4 tetraphenylcyclobutane, space group $P2_1/a$ (Dunitz, 1949).

(2) Acentric distribution. General reflexions from nitronium perchlorate (Truter, 1950). The space group of this substance was found in an earlier statistical investigation to be Cc .

Each intensity, I , observed by these workers was replaced by I_o , where

$$I_o = I(1 \pm n\delta). \quad (14)$$

The sign and the magnitude of the error, δ , for each reflexion were decided using the tables of random numbers given by Fisher & Yates (1948), the actual value of δ being determined from tables of the normal error function.

The adjustable factor, n , was introduced to control the severity of the errors introduced; the three values used gave r.m.s. errors, σ , of 0.15, 0.36 and 0.70.*

* For severe errors I_o (in equation (14)) became negative, but in such cases the intensities were reckoned as zero, so

2.3.2. *The N(z) test.*—The $N_o(z)$ distribution curves are shown in Fig. 1(a) and (b). It can be seen that

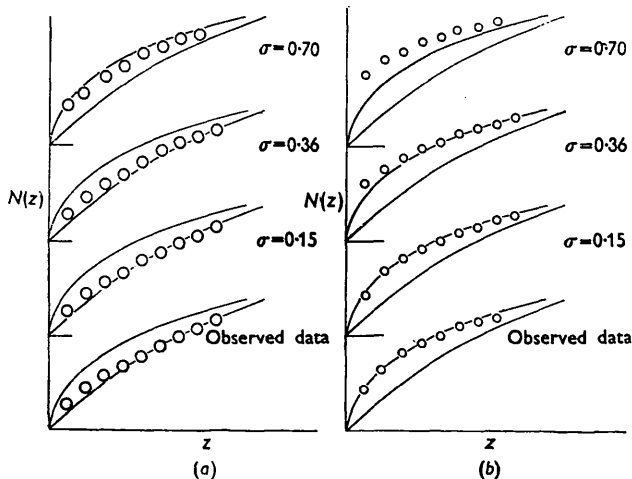


Fig. 1. The modification of the cumulative distribution, $N(z)$, resulting from the introduction of random errors. σ is the R.M.S. fractional error, and is defined in § 2.2.

- (a) $N(z)$ for an acentric distribution of intensities.
- (b) $N(z)$ for a centric distribution of intensities.

even for quite large errors ($\sigma \sim 0.4$) the distributions are only slightly distorted, and it is not until very large values of σ (~ 0.7) are introduced that there is sufficient distortion to lead to a false interpretation. Such large errors cause an acentric distribution to be distorted to such an extent that it would better fit, and thus be mistaken for, the centric distribution; no such confusion would arise in the present investigation until σ has values approaching 0.7. Similarly, the distortion of the centric curve is not likely to lead to confusion with the hypercentric (Lipson & Woolfson, 1952; Rogers & Wilson, 1953) so long as random errors are the sole cause of the distortion.

2.3.3. *The variance and ratio tests.*—The values obtained experimentally for ϱ_o and v_o are set out in Tables 2 and 3. From them and from equations (11) and (12) we can draw the following conclusions, bearing in mind, however, that these equations presuppose an undistorted normal distribution of errors:

- (1) An acentric set of intensities will not be mistaken for centric by either test unless random errors occur for which σ lies between 0.36 and 0.70 and rather nearer the latter.
- (2) It is a curious fact, verifiable by reference to

that some departure from a strictly Gaussian distribution of errors then occurs, and the relation $\sigma = (\pi/2)^{1/2} \langle |\Delta| \rangle$ will not apply for the larger values of σ ($\gtrsim \frac{1}{2}$) treated here. The values of σ quoted here are for the errors actually introduced, i.e.

$$\sigma = \langle \Delta^2 \rangle^{1/2} = \langle n^2 \delta^2 \rangle^{1/2},$$

and it is only for the lower values of n that one can use the approximation

$$\sigma = n \langle \delta^2 \rangle^{1/2}.$$

Table 2

σ of random errors introduced	Acentric			Centric		
	v_c	v_1	v_2	v_c	v_1	v_2
0.00	1.000	1.04	1.04	2.000	2.07	2.07
0.15	1.045	1.08	1.08	2.068	2.02	2.02
0.36	1.259	1.25	1.24	2.388	2.38	2.38
0.70	1.980	1.98	1.74	3.471	3.51	2.68

$v_1 = \langle z'^2 \rangle - 1$ where $z' = I_o / \langle I \rangle$, and $v_2 = \langle z_o^2 \rangle - 1$, where $z_o = I_o / \langle I_o \rangle$.

The difference between $\langle I \rangle$ and $\langle I_o \rangle$ is a consequence of the asymmetry referred to in the footnote, p. 385 (second column). It is noteworthy that v_2 , which would usually be evaluated, is less affected by random errors than v_1 .

Table 3

σ of random errors introduced	Acentric		Centric	
	ϱ_1	ϱ_2	ϱ_1	ϱ_2
0.00	0.774	0.774	0.656	0.656
0.15	0.768	0.768	0.639	0.639
0.36	0.754	0.743	0.626	0.613
0.70	0.684	0.674	0.569	0.562

ϱ_1 and ϱ_2 correspond to v_1 and v_2 of Table 2.

equations (34) and (35) of Rogers & Wilson (1953), that each of the distributions in the sequences, acentric, centric, bicentric ... or acentric (= aparallel), parallel, biparallel ... has a specific variance, v_n , which would be increased by random errors to that of the next higher member, v_{n+1} , for $\sigma^2 = 0.5$ or $\sigma = 0.707$. The ratio $\varrho_n / \varrho_{n+1}$ is likewise a constant, viz. $\pi^2/8$, but it is not possible to estimate the corresponding value of σ as the series (12b) is divergent.

(3) Similarly, the value $\frac{1}{2}(v_n + v_{n+1})$ always corresponds to $\sigma = 0.50$. This is roughly, therefore, the upper limit of σ for which correct interpretations can be made.

(4) The values of v and ϱ prior to our introduction of random errors contain both the effects of the experimental errors of observation by the original investigators and the errors involved in the statistical manipulation. For the variances their combined effects can be interpreted as $\sigma = 0.14$ (acentric) and 0.15 (centric), but it seems unwise to place much reliance on a value of σ deduced in this way.

2.4. The average-intensity multiple

It is evident that if the errors are symmetrically distributed ($\langle \Delta \rangle = 0$) the mean value of I is unchanged, and this will be true of both the enhanced row or zone and the others used for comparison. The multiple found is, therefore, free from systematic error so long as symmetrically distributed random errors only occur.

2.5. According to Robertson & White (1947), the mean random error to be expected in photographic work is not likely to exceed 20% in I , i.e. $\langle |\Delta| \rangle = 0.2$, or, assuming the errors to be normally distributed, $\sigma = 0.25$, $\sigma^2 = 0.06$.

Random errors of this magnitude are not able, therefore, to lead to misinterpretation of the results given by any of the statistical tests.

3. Errors systematic as a function of I only

3.1. Systematic errors which are a function of the intensity only may arise from the methods of recording and estimating the intensities (for further details see Kaan & Cole (1949) and Wallwork & Standley (1954)), from inadequate allowance for lost counts with Geiger-counter equipment, and from extinction. We shall defer consideration of the unobserved reflexions to § 5; for the rest the most probable distortion of the intensities would seem to be underestimation of the strong intensities and overestimation of the weaker. Errors which compress the range of intensities in this way will reduce the dispersion. Conversely, errors which expand the intensity range increase the dispersion. This is in contrast to the effects produced by random errors, which can only increase the dispersion.

A simple type of compressive distortion, which is mathematically convenient for estimating the magnitude of the effects on the statistical results, and is also a satisfactory first-order approximation to the distortions just described, is obtained by supposing a relation*

$$I_o = k\Sigma\{1 - \exp(-I/k\Sigma)\}, \quad (15)$$

where $\Sigma = \langle I \rangle$. If we assume that 10% underestimation for $I = 3\langle I \rangle$ is reasonably severe we find that $k \approx 15$, and will normally exceed this.

In addition to z and z_o , whose averages are unity, we shall need to use

$$z' = I_o/\langle I \rangle = g_1 z_o, \quad (16)$$

whose average is g_1 . Rewriting equation (15) gives

$$z' = k\{1 - \exp(-z/k)\}, \quad (17)$$

so that

$$0 \leq z' \leq k$$

and

$$z = -k \ln(1 - z'/k). \quad (18)$$

We shall determine first $\langle z' \rangle = g_1$ and then $N_o(z)$ in each case.

* Another distortion of the form

$$I_o = k\Sigma \tanh(I/k\Sigma)$$

has also been considered because of its close connection with Darwin's (1922) result for primary extinction. (This result, however, applies only to large crystal slabs with the reflecting planes parallel to the external face; see also Weiss, 1952.) This function was used to obtain analytic results for the acentric case which were of much the same order of magnitude as those quoted here for the exponential function. The centric case led, however, to a very slowly converging series quite unsuited to numerical computation.

3.2. The $N(z)$ test

3.2.1. *Acentric case.*—The ideal acentric distribution is

$$(1)P(z)dz = \exp(-z)dz. \quad (19)$$

Hence,

$$(1)P(z')dz' = (1 - z'/k)^{k-1}dz'$$

and

$$(1)g_1 = \langle z' \rangle = \int_0^k (1 - z'/k)^{k-1} z' dz' \quad (20)$$

$$= k/(k+1) \quad (21)$$

$$= 1 - k^{-1} + k^{-2} - \dots \quad (21a)$$

Thus

$$z_o = z'(k+1)/k = (k+1)\{1 - \exp(-z/k)\}. \quad (22)$$

For low values of z/k the change to z_o is due entirely to the reduced average and is equivalent to a linear expansion of the scale of abscissae in the $(1)P(z)$ and $(1)N(z)$ graphs. More generally, the whole $(1)N_o(z)$ curve can be obtained by plotting the usual ordinates of the acentric $N(z)$ graph against the transformed abscissae z_o . Alternatively, from the definition

$$(1)N_o(z) = \int_0^{z_o} P(z_o) dz_o,$$

we get

$$(1)N_o(z) = \{1 - [1 - z_o/(1+k)]^k\}, \quad (23)$$

whereas

$$(1)N(z) = \{1 - \exp(-z)\}. \quad (24)$$

Equation (23) is, however, less convenient for plotting the $(1)N_o(z)$ graph. Fig. 2(a) shows that the departure is slight for $k > 10$ and tends away from the centric curve.

To estimate k , it is convenient to plot $(1)N_o(z)/(1)N(z)$ against z . The result (in the range $0 < z < 1$) is very nearly a straight line with an intercept $(1)g_1 = k/(k+1)$ and a slight positive slope. If the experimental values of $(1)N_o(z)$ are sufficiently consistent it may be possible to estimate k from this intercept.

3.2.2. *Centric case.*—Starting from the centric distribution

$$(\bar{1})P(z)dz = (2\pi z)^{-\frac{1}{2}} \exp(-z/2)dz, \quad (25)$$

and following a similar procedure gives

$$(\bar{1})g_1 = \int_0^k \frac{(1 - z'/k)^{\frac{1}{2}k-1} z' dz'}{\{2\pi k \ln(1 - z'/k)\}^{\frac{1}{2}}}, \quad (26)$$

which is conveniently integrated by substituting

$$(1 - z'/k) = \exp(-p^2).$$

As a result

$$(\bar{1})g_1 = k\{1 - [k/(k+2)]^{\frac{1}{2}}\} \quad (27)$$

$$= 1 - (3/2k) + (5/2k^2) - \dots \quad (27a)$$

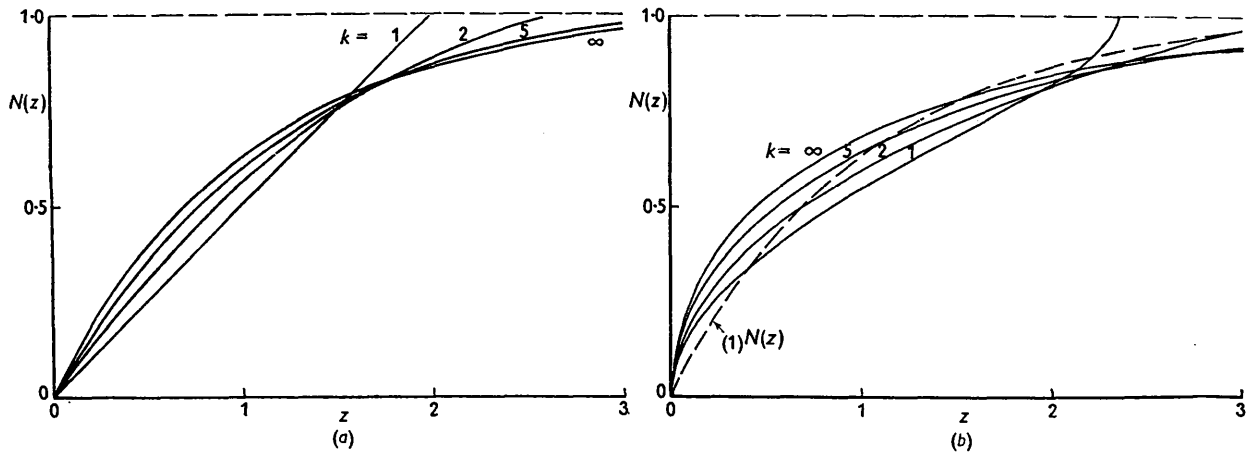


Fig. 2. The modification of the cumulative distribution, $N(z)$, resulting from the introduction of a systematic 'exponential' error as defined in § 3.1. (a) $N(z)$ for a set of acentric intensities. (b) $N(z)$ for a set of centric intensities.

(This is a 50% larger change than for $(1)g_1$ at the usual large values of k .) Hence,

$$z_0 = \frac{1 - \exp(-z/k)}{1 - [k/(k+2)]^{\frac{1}{2}}} \quad (28)$$

and

$$(\bar{1})N_o(z) = \text{erf} \left[-\frac{1}{2}k \ln \{1 - z_0(1 - [k/(k+2)]^{\frac{1}{2}})\} \right]^{\frac{1}{2}}, \quad (29)$$

which, for small values of z/k , reduces to

$$\text{erf} \left[\frac{1}{2}g_1 z_0 \right]^{\frac{1}{2}}, \quad (29a)$$

whereas

$$(\bar{1})N(z) = \text{erf} \left[\frac{1}{2}z \right]^{\frac{1}{2}}. \quad (30)$$

Here, too, it is much easier to plot $(\bar{1})N_o(z)$ by transforming the abscissae only, a transformation which is again linear for low values of z/k . Fig. 2(b) indicates the trend of $(\bar{1})N_o(z)$, and it is evident that for $k \geq 5$ there will be no confusion with the acentric distribution. The magnitude of k can be estimated by plotting either

(a) $(\bar{1})N_o(z)/(\bar{1})N(z)$ against z , or

(b) z_0/z against z for $(\bar{1})N(z_0) = (\bar{1})N(z)$ (see footnote*, p. 384).

These are both nearly linear in the range $0 < z < 1$. The first has an intercept $[(\bar{1})g_1]^{\frac{1}{2}}$, the second an intercept $(\bar{1})g_1$.

3.3. The variance test

3.3.1. *Acentric case.*—By analogy with equation (20) we have

$$(1)g_2 = \langle (z')^2 \rangle = \int_0^k (1 - z'/k)^{k-1} (z')^2 dz' \quad (31)$$

$$= 2k^2/(k+1)(k+2), \quad (32)$$

so that

$$(1)v_o = k/(k+2). \quad (33)$$

For $k \rightarrow \infty$ this becomes unity, as found by Wilson,

and it is evident that it will not be seriously affected even by values of k as low as 10.

3.3.2. *Centric case.*—Analogy with equation (26) and procedure as in § 3.3.1 gives

$$(\bar{1})g_2 = \langle (z')^2 \rangle = k^2 \left[1 - 2 \left(\frac{k}{k+2} \right)^{\frac{1}{2}} + \left(\frac{k}{k+4} \right)^{\frac{1}{2}} \right], \quad (34)$$

so that

$$(\bar{1})v_o = \frac{\left(\frac{k}{k+4} \right)^{\frac{1}{2}} - \left(\frac{k}{k+2} \right)^{\frac{1}{2}}}{\left[1 - \left(\frac{k}{k+2} \right)^{\frac{1}{2}} \right]^2}. \quad (35)$$

On expansion this gives

$$(\bar{1})v_o = 2 \left[1 - \frac{3}{k} + \frac{47}{4k^2} - \frac{587}{4k^3} + \dots \right], \quad (35a)$$

which again agrees with Wilson's result as $k \rightarrow \infty$.

Equation (35) gives

$$(\bar{1})v_o = 1.661 \quad \text{for } k = 15$$

and

$$(\bar{1})v_o = 1.295 \quad \text{for } k = 5,$$

so that there is little likelihood of confusion occurring in practice.

3.4. The test ratio

3.4.1. *Acentric case.*—It is easily shown that

$$(1)g_{\frac{3}{2}} = k^{\frac{3}{2}} \frac{\Gamma(k)\Gamma(\frac{3}{2})}{\Gamma(k+\frac{3}{2})}, \quad (36)$$

whence

$$(1)g_o = \frac{\pi}{4} \left\{ \frac{\Gamma(k+1)\Gamma(k+2)}{\Gamma^2(k+\frac{3}{2})} \right\}. \quad (37)$$

This may be reduced by use of Stirling's formula and represented to a very close approximation (for $k \geq 5$) by

$$\frac{\pi}{4} \exp\left(\frac{1}{4k}\right). \quad (38)$$

As $k \rightarrow \infty$ the quantity in the braces of equation (37) tends to unity so that $(1)\rho_o \rightarrow \pi/4 = 0.785$, as found by Wilson. At $k = 15$, $(1)\rho_o = 0.798$; and at $k = 5$, $(1)\rho_o = 0.819$.

3.4.2. *Centric case.*—Following the pattern of equations (26) and (34) we obtain

$$(\bar{1})g_{\frac{1}{2}} = k \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty \exp(-\frac{1}{2}kp^2) [1 - \exp(-p^2)]^{\frac{1}{2}} dp, \quad (39)$$

which is most conveniently evaluated by expanding the surd as a polynomial in p . This gives

$$(\bar{1})g_{\frac{1}{2}} = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left\{1 - \frac{1}{2k} + \frac{5}{12k^2} + \dots\right\}, \quad (40)$$

whence

$$(\bar{1})\rho_o = \frac{2}{\pi} \cdot \frac{\left(1 - \frac{1}{2k} + \frac{5}{12k^2} \dots\right)^2}{k \left[1 - \left(\frac{k}{k+2}\right)^{\frac{1}{2}}\right]}. \quad (41)$$

For $k = 15$ this gives $(2/\pi) \times 1.029 = 0.655$ and for $k \rightarrow \infty$, $(\bar{1})\rho_o \rightarrow 2/\pi = 0.637$, the figure given by Wilson.

The increment here is quite insignificant compared with the interval between $2/\pi$ (centric) = 0.637 and $\pi/4$ (acentric) = 0.785 and there can be no possibility of confusion arising from this type of error when the ratio test is used to distinguish between the distributions.

3.5. The average-intensity multiples

For a set of reflexions with enhanced $\langle I \rangle$ the probabilities of equations (19) and (25) are replaced by

$$(1)P(I)dI = (n\Sigma)^{-1} \exp(-I/n\Sigma)dI \quad (42)$$

and

$$(\bar{1})P(I)dI = (2\pi n\Sigma I)^{-\frac{1}{2}} \exp(-I/2n\Sigma)dI, \quad (43)$$

where n is the multiple sought (see Wilson, 1951).

Equation (42) gives, in place of equation (20),

$$[(1)g_1]_n = n^{-1} \int_0^k (1-z'/k)^{\frac{k}{n}-1} z' dz' \quad (44)$$

$$= nk/(k+n). \quad (45)$$

Similarly equation (26) becomes

$$[(\bar{1})g_1]_n = \int_0^k \frac{(1-z'/k)^{\frac{k}{2n}-1} z' dz'}{2\pi nk \ln(1-z'/k)^{-1}} \quad (46)$$

$$= k \left\{1 - \left(\frac{k}{k+2n}\right)^{\frac{1}{2}}\right\}. \quad (47)$$

The multiple obtained as the ratio for two zones, one normal and the other enhanced, is, therefore,

$$n_o = [g_1]_n/[g_1]_1. \quad (48)$$

For two acentric zones

$$n_o = n(k+1)/(k+n), \quad (49)$$

which, as $k \rightarrow \infty$, tends to n , as it should. For $n = 2$ (a plane of symmetry) and $k = 10$, n_o drops only to 1.83, and for $k = 5$ to 1.72. For two centric zones

$$n_o = \frac{1 - [k/(k+2n)]^{\frac{1}{2}}}{1 - [k/(k+2)]^{\frac{1}{2}}} \quad (50)$$

$$= n \left\{1 - \frac{3(n-1)}{2k} + \dots\right\}, \quad (51)$$

which also tends to n as $k \rightarrow \infty$. Here, for $n = 2$, $k = 10$, the observed multiple is 1.77, and for $k = 5$ is 1.65. The comparison of a centric and an acentric follows similar lines.

These results indicate, as might have been expected, that the more disperse centric zones will be the more affected. Nevertheless, so long as this is the sole source of error the issue is never likely to be in doubt in normal problems.

4. Errors systematic as a function of $s = (2 \sin \theta)/\lambda$

4.1. Errors of this type may occur in the estimated intensities as a result of changes in the spot size and shape, in the α -doublet separation and in the background density. They may also arise in the estimation of absorption corrections. It was pointed out, however, by Howells *et al.* (1950) that if ideally performed the statistical manipulations should yield results unaffected by errors of this kind. For, as was shown by these authors, the statistical tests are most conveniently carried out on the data expressed in terms of $z (= I/\langle I \rangle)$, which will not contain errors systematic in s , if the course of the $\langle I \rangle$ curve has been accurately determined. The remainder of this section is devoted, therefore, to a study of the effects of using erroneous values of $\langle I \rangle$ for preparing the z data.

Before doing so, however, it is necessary to emphasize that these errors are not eliminated in the method proposed by Wilson (1942) and now widely used for putting the observed intensities on an approximately absolute scale. Their effects are considered in § 4.4, but it is worth noting here that both the scaling factor and the exponent in the 'temperature factor' are affected. The error in the scaling factor from such causes will usually be small, but the error in the exponent of the temperature factor may be large and it is then no longer properly identifiable with the Debye-Waller constant.

4.2. Effect on the distribution functions

If we suppose that the intensities have been accurately estimated, but that the value determined for the local average, $\langle I \rangle_{\text{det.}}$, is incorrect, we may write,

$$\langle I \rangle_{\text{det.}} = \langle I \rangle [1 + \varepsilon(s)], \quad (52)$$

where $\varepsilon(s)$ is a function of s , which, one hopes, is never large compared with unity. Then,

$$z_{\text{det.}} = z/[1 + \varepsilon(s)] \quad (53)$$

and the empirical cumulative distribution determined from all reflexions in a small range of s will be, aside from random errors and statistical fluctuations,

$$N_o(z_{\text{det.}}) = N\{z_{\text{det.}}[1 + \varepsilon(s)]\}, \quad (54)$$

where $N(z)$ is the ideal cumulative distribution function. The observed distribution function, $N_o(z)$ (see footnote *, p. 384), is the average of those obtained for the range of s considered.* This is usually the whole range for which reflexions are observed, but sometimes the low-angle and high-angle reflexions are considered separately (see § 4.4).

We may then expand equation (54) and obtain

$$\begin{aligned} N_o(z) &= \langle N(z) + \varepsilon(s)zP(z) + \frac{1}{2}\varepsilon^2(s)z^2P'(z) + \dots \rangle \\ &= N(z) + \langle \varepsilon \rangle zP(z) + \frac{1}{2}\langle \varepsilon^2 \rangle z^2P'(z) + \dots \end{aligned} \quad (55)$$

The probability distribution of z is the derivative of this,

$$P_o(z) = (1 + \langle \varepsilon \rangle)P(z) + (\langle \varepsilon \rangle + \langle \varepsilon^2 \rangle)zP'(z) + \dots, \quad (56)$$

where $P(z)$ is the ideal distribution function. Thus the first-order effect on both $N(z)$ and $P(z)$ can be either an enhancement or a depression, depending on the sign of $\langle \varepsilon \rangle$.

4.3. Effect on the variance and ratio tests

From equation (53) the observed mean value of $z_{\text{det.}}$ can be written as

$$\langle z_{\text{det.}} \rangle = \langle z \rangle_o = \langle \langle z\psi(s) \rangle \rangle, \quad (57)$$

where one average is with respect to z , the other is with respect to s , and

$$\psi(s) = \{1 + \varepsilon(s)\}^{-1}. \quad (58)$$

Since $\langle z \rangle$ is unity for all distributions (Rogers & Wilson, 1953), equation (57) reduces to

$$\langle z \rangle_o = \langle \psi(s) \rangle. \quad (59)$$

Similarly, the mean value of $z_{\text{det.}}^2$ is

$$\langle z^2 \rangle_o = \langle \langle z^2 \psi^2(s) \rangle \rangle \quad (60)$$

$$= (v+1)\langle \psi^2(s) \rangle, \quad (61)$$

* Note that in forming the average of a function, $f(s)$, different weights are appropriate for rows, zones and complete arrays:

$$\begin{aligned} \langle f \rangle &= \frac{1}{s_2 - s_1} \int_{s_1}^{s_2} f(s) ds && \text{(row)}, \\ &= \frac{2}{s_2^2 - s_1^2} \int_{s_1}^{s_2} f(s) s ds && \text{(zone)}, \\ &= \frac{3}{s_2^3 - s_1^3} \int_{s_1}^{s_2} f(s) s^2 ds && \text{(array)}. \end{aligned}$$

where v is the specific variance. The observed variance may take one of three forms, depending on exactly how it is evaluated.

If the mean value of $z_{\text{det.}}$ is assumed to be unity, as it would be in the absence of error, the determined value of the specific variance is

$$\begin{aligned} \text{(i)} \quad v_o &= \langle z^2 \rangle_o - 1 \\ &= v + (v+1)\{\langle \psi^2(s) \rangle - 1\}, \end{aligned} \quad (62)$$

which is exactly analogous to equation (12).

If the actual mean value of $z_{\text{det.}}$ for the whole range of s is used

$$\text{(ii)} \quad v_o = v\langle \psi^2(s) \rangle + \langle \psi^2(s) \rangle - \langle \psi(s) \rangle^2; \quad (63)$$

the terms not multiplied by v are the variance of $\psi(s)$.

Finally, the square of the local mean value of $z_{\text{det.}}$, $\psi^2(s)$, could be subtracted from the local mean value of $z_{\text{det.}}^2$, $(v+1)\psi^2(s)$, before the averaging with respect to s is carried out. The observed specific variance then has the value

$$\text{(iii)} \quad v_o = v\langle \psi^2(s) \rangle. \quad (64)$$

These three results are conveniently compared if written in the form

$$\begin{aligned} \text{(i)} \quad v_o &= v + (v+1)\langle \psi^2(s) - 1 \rangle \\ &= v\langle \psi^2(s) \rangle + \langle \psi^2(s) \rangle - 1, \end{aligned} \quad (62a)$$

$$\begin{aligned} \text{(ii)} \quad v_o &= v + (v+1)\langle \psi^2(s) - 1 \rangle - \langle \psi(s) \rangle^2 + 1 \\ &= v\langle \psi^2(s) \rangle + \langle \psi^2(s) \rangle - \langle \psi(s) \rangle^2, \end{aligned} \quad (63a)$$

$$\begin{aligned} \text{(iii)} \quad v_o &= v + v\langle \psi^2(s) - 1 \rangle \\ &= v\langle \psi^2(s) \rangle. \end{aligned} \quad (64a)$$

Of the three, the third seems preferable, although the second has been that ordinarily used in the past as it involves less work.

The observed values of the test ratio will be

$$\varrho_o = \langle \langle z^2 \psi^2(s) \rangle \rangle_s, \quad (65)$$

the suffixes indicating the averaging variables. This transforms to

$$\varrho_o = \varrho\langle \psi(s) \rangle \quad \text{or} \quad \varrho\langle z \rangle_o. \quad (66)$$

The observed values of the specific variance or test ratio, like the distribution functions, may, therefore, be either reduced or increased, depending on the sign of $\langle \varepsilon \rangle$.

4.4. The experimentally determined $\langle I \rangle$ curve usually corresponds to an equation of the form

$$\langle I \rangle_{\text{det.}} = A \Sigma \exp \left\{ -\frac{1}{2} B s^2 \right\}, \quad (67)$$

where $\Sigma = \Sigma f_i^2$ and may be calculated from tabulated values of the atomic scattering factors, and A and B are the scaling factor and the effective temperature coefficient respectively. A , B are determined experimentally from a plot of $\log \{ \langle I \rangle_{\text{det.}} / \Sigma \}$ versus s^2 or $\sin^2 \theta$ by drawing the best straight line through

the points. Then, assuming the validity of equation (67), the intercept gives A and the gradient B . The values estimated in this way may be in error because of: (i) the use of inappropriate scattering factors (see Rogers, 1954, 1955), (ii) incorrect allowances for absorption in spherical or cylindrical crystals, (iii) an unknown amount of overlapping of atoms in projection, which raises $\langle I \rangle$ for small values of s , (iv) extinction, which reduces $\langle I \rangle$ for s small, and (v) the effects of unobserved reflexions (see § 5.1).

These can all impart curvature to the $\log \{ \langle I \rangle_{\text{det.}} / \Sigma \}$ plot, which, when combined with the inevitable scatter of the points, raises some doubt as to the best straight line and what weight to attach to the various points. (Some examples are discussed by Rogers (1954, 1955) in which the curvature is sufficiently well defined to justify the abandonment of a straight line.)

If, however, we write

$$\langle I \rangle_{\text{det.}} = A(1 + \alpha)\Sigma(1 + \sigma) \exp \left\{ -\frac{1}{2}(B + \beta)s^2 \right\}, \quad (68)$$

where A , B , Σ are the correct values and α , β , σ are their errors, it is evident that only σ can be a function of s . Then, from (52),

$$\begin{aligned} 1 + \varepsilon(s) &= (1 + \alpha)(1 + \sigma) \exp \left\{ -\frac{1}{2}\beta s^2 \right\}, \\ \varepsilon(s) &= \alpha + \sigma - \frac{1}{2}\beta s^2, \end{aligned} \quad (69)$$

so that $\langle \varepsilon \rangle$, which determines to a first approximation the departure for each test, becomes

$$\langle \varepsilon \rangle = \alpha + \langle \sigma \rangle - \frac{1}{2}\beta \langle s^2 \rangle. \quad (70)$$

(As explained previously, the values of $\langle s^2 \rangle$ depend on whether a row, zone or array is being considered.) Errors in the intercept affect the low- and high-angle values of $\langle I \rangle$ equally. Errors in the slope will affect the high-angle reflexions preferentially. In fact, with suitable signs for α , β , σ the $N(z)$ curve (or other statistical test) may show opposite departures from the ideal at high and low angles. These effects are, however, confined to the statistical tests and the reliability index, since errors in converting to the absolute scale of intensities can have little influence on the elucidation of a structure and can be removed at a later stage by adjustments designed to make $\langle I_o \rangle / \langle I_t \rangle$ or $\langle |F_o| \rangle / \langle |F_t| \rangle$ unity for all values of s .

4.5. Possibilities of confusion in the tests for centrosymmetry

For both the acentric and the centric distributions the quantity $zP(z)$, appearing as the coefficient of $\langle \varepsilon \rangle$ in (55) and (56), has its maximum values (0.368 and 0.242 respectively) at $z = 1$. Here the difference between the cumulative distribution functions is only 5%, so that errors of as little as 15% in the $\langle I \rangle$ curve might lead to some hesitation as to the type of distribution. However, the shapes of the distributions at lower values of z are very little affected, and a reconsideration of the $\langle I \rangle$ curve might serve to remove the hesitation. Much larger errors, approaching 100%,

would be needed to produce confusion of the distribution at $z \sim 0.3$. To raise the value of the specific variance from 1 to 1.5 would need a mean error of about -20% and to drop it from 2 to 1.5 would need about +15%. Errors of 10% would suffice to make the test ratio indecisive.

It thus appears that errors in the $\langle I \rangle$ curve are more likely to lead to hesitation or even mistakes in deciding on the type of distribution than are random errors or systematic errors of the extinction type. For the successful application of the statistical tests it is essential, therefore, to determine the $\langle I \rangle$ function as accurately as possible. Recommendations for this are given by Rogers (1955).

4.6. The average-intensity multiples

The values of $\langle I \rangle$ found in two zones at the same $\langle \sin^2 \theta \rangle$ will differ if there is anisotropy of the thermal motion in the crystal or if the factors (ii), (iii), (iv) or (v) of § 4.4 have different disturbing effects in the two zones. The graphs of $\log \langle I \rangle$ versus $\langle \sin^2 \theta \rangle$ will not then coincide. Similarly, if one zone has an enhanced average the two curves will not run parallel so that the ratio $\langle I \rangle_n / \langle I \rangle$ proves to be a function of θ . The increase due to atomic overlap and the decrease due to extinction are both confined to low θ , whereas threshold effects (see § 5.1) are then least. Provided, therefore, that the residual absorption errors are small, the high-angle parts of the $\log \langle I \rangle$ versus $\langle \sin^2 \theta \rangle$ graphs should taper uniformly. It is from these portions that extrapolations back to $\theta = 0$ should be made in order to eliminate the effects of the differing temperature factors and threshold effects. Although the precision of the intercepts may be small it should be adequate in view of the large effect sought.

5. Errors due to unobserved reflexions

5.1. Most workers are in the habit of estimating the lowest observable intensity that can be measured in different parts of their films. This varies only slightly with $\sin \theta$, but after allowing for the Lorentz and polarization factors the threshold intensity, I_t , varies considerably with $\sin \theta$. Intensities of finite magnitude less than this are either not observed at all or are too faint for any reliable estimate to be made. Clearly I_t is not directly dependent on the intensity distribution; indirectly it may be influenced by the investigator's decision to use larger exposures for centrosymmetric crystals. We are assuming, therefore, a relation between I_o and I of the form

$$\left. \begin{aligned} I_o &= I & \text{for } I \geq I_t, \\ &= 0 & \text{for } I < I_t. \end{aligned} \right\} \quad (71)$$

This gives, on reference to the two distribution functions,

$$(1)g_1 = (z_t + 1) \exp(-z_t) \quad (72)$$

and

$$(\bar{I})g_1 = 1 - \operatorname{erf}(\frac{1}{2}z_t) + (2z_t/\pi)^{\frac{1}{2}} \exp(-\frac{1}{2}z_t), \quad (73)$$

where

$$z_t = I_t/\Sigma \quad \text{or} \quad I_t/\langle I \rangle.$$

These functions are depicted in Fig. 3(a).

For many organic compounds the attenuation with $\sin^2 \theta$ is so rapid that few reflexions are observable at high θ . The value then obtained for $\langle I_o \rangle$ will be very much lower than $\langle I \rangle$ and this will be a smooth factor of major importance contributing to the effective temperature coefficient. The broken lines in

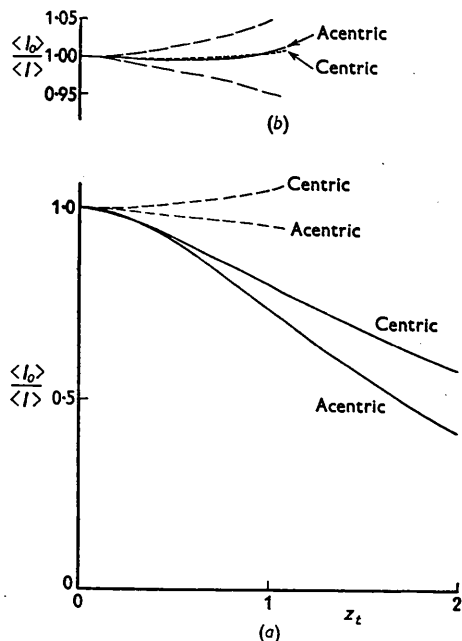


Fig. 3. (a) The ratio $\langle I_o \rangle / \langle I \rangle$ which results if all unobserved reflexions ($I \lesssim z_t \Sigma$) are included in the averaging with zero intensity (full lines), or $0.35z_t \Sigma$ (broken lines).

(b) Improvement obtained by including unobserved reflexions at $0.43z_t \Sigma$ (acentric) or $0.30z_t \Sigma$ (centric). Broken lines as in (a). Note the change in scale of the ordinates.

Fig. 3(a), (b) illustrate the improvements which result from including all these unobserved reflexions at $0.35z_t$ when finding $\langle I_o \rangle$. If, however, the distribution of intensities is already known the best approximations are given by $0.43z_t$ (acentric) and $0.30z_t$ (centric), (see Fig. 3(b)). The influence of unobserved reflexions on the reliability index has already been discussed by Phillips, Rogers & Wilson (1950), and Hamilton (1955) has discussed their proper use in least-squares refinement.

5.2. The $N(z)$ test

From equation (71) we have

$$\begin{aligned} z_o &= I_o / \langle I_o \rangle = I / \langle I_o \rangle, \\ &= z/g_1 \quad \text{for } I \geq I_t \\ &= 0 \quad \text{for } I \leq I_t. \end{aligned} \quad (74)$$

and

The graph of $N_o(z)$ is obtained, therefore, by re-scaling the z axis for all values of $z_o \geq z_t/g_1$, but is constant below this point (see Fig. 4). The sharp

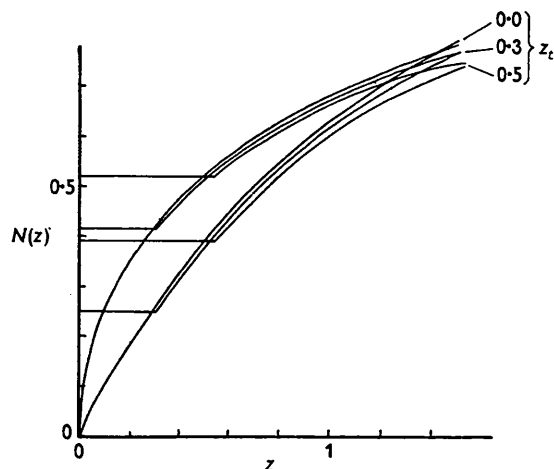


Fig. 4. $N(z)$ curves as modified by including all unobserved reflexions with zero intensity for threshold intensities, z_t , of 0.0, 0.3, 0.5.

break in these curves will be rounded in practice as a result of combining data for which z_t is changing rapidly in reciprocal space. Again, the most satisfactory parts of these plots will be at $z \sim 0.3$ to 0.5 and parts of the reciprocal lattice which contain more than about 50% (centric) or 30% (acentric) of unobserved reflexions should preferably be omitted.

5.3. The variance test

It is easy to show that

$$(1)g_2 = \frac{1}{2}(z_t^2 + 2z_t + 2) \exp(-z_t) \quad (75)$$

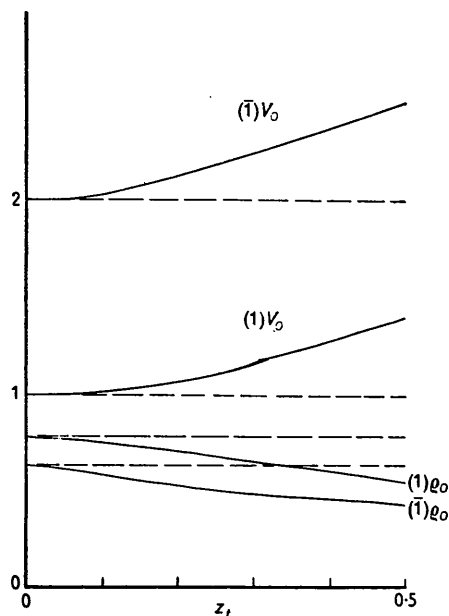


Fig. 5. The dependence of the specific variances and the test ratios upon the threshold intensity, z_t .

and

$$(\bar{1})g_2 = (\bar{1})g_1 + \frac{1}{3}(2z_i^3/\pi)^{\frac{1}{2}} \exp(-z_i/2). \quad (76)$$

Hence, from (7),

$$(1)v_o = \exp(z_i) \{1 + (z_i + 1)^{-2}\} - 1 \quad (77)$$

and

$$(\bar{1})v_o = \frac{3}{(\bar{1})g_1} + \left(\frac{2z_i^3}{\pi}\right)^{\frac{1}{2}} \cdot \frac{\exp(-\frac{1}{2}z_i)}{[(\bar{1})g_1]^2} - 1, \quad (78)$$

which are plotted in Fig. 5. There is little possibility of confusion for z_i up to 0.5.

5.4. The ratio test

The quantities, $g_{\frac{1}{2}}$, may be shown to be

$$(1)g_{\frac{1}{2}} = 2(z_i/\pi)^{\frac{1}{2}} \exp(-z_i) + (1 - \operatorname{erf}(z_i^{\frac{1}{2}})) \quad (79)$$

and

$$(\bar{1})g_{\frac{1}{2}} = \exp(-\frac{1}{2}z_i), \quad (80)$$

so that, by (5),

$$(1)\rho_o = (1 + z_i) \{(\pi/4)(1 - \operatorname{erf}(z_i^{\frac{1}{2}}))^2 \exp(z_i) + (\pi z_i)^{\frac{1}{2}}(1 - \operatorname{erf}(z_i^{\frac{1}{2}})) + z_i \exp(-z_i)\} \quad (81)$$

and

$$(\bar{1})\rho_o = 2 \exp(-z_i)/\pi(\bar{1})g_1. \quad (82)$$

From Fig. 5, in which these are plotted, it is immediately apparent that the ratio test is very sensitive to errors of this kind and confusion will occur for $z_i \gtrsim 0.15$.

5.5. Average multiples

The effects of unobserved reflexions on the average-intensity multiples have been considered already in § 4.4.

6. Conclusions

We can now attempt to review the detailed conclusions set out in the several sections of this paper.

(1) Neither random errors, nor systematic errors of the 'extinction' type, having ordinary magnitudes, can lead to ambiguity or confusion of the distribution types. In particular, where these errors coexist, the deviations they each produce are opposed and thus tend to cancel.

(2) Special care must be taken in the evaluation of the $\langle I \rangle$ function, and those annuli for which $I_i \gtrsim \frac{1}{3}\langle I_o \rangle$ should either be discarded or used with caution. For further details see the recommendations given by Rogers (1955).

(3) The three tests for centrosymmetry are unequally affected by any given error. The variance is most disturbed by a redistribution of the strong terms, the ratio test by redistribution of the weak terms. The $N(z)$ test, however, keeps these distinct and is to that extent more informative. In serious statistical work

it seems desirable to use all three tests side by side and if possible to separate the results for different ranges of $\sin^2 \theta$.

(4) Although the value deduced for the average-intensity multiple may not be very accurate, the identification of a plane of symmetry should never be in doubt as the effect sought is so large.

(5) The results derived here permit the estimation of a likely outer limit for departures from the ideal statistical criteria in any problem. Anomalies exceeding these limits can be regarded as significant and an explanation should be sought on the lines indicated in (i), (ii), (iii) of § 1.1.

One of the authors (D. R.) is indebted to the University of Wales for an I.C.I. Research Fellowship during the tenure of which this work was completed. Another (E. S.) acknowledges gratefully the financial assistance of the University of Wales and the Department of Scientific and Industrial Research.

References

- BOOTH, A. D. (1947). *Proc. Roy. Soc. A*, **188**, 77.
 BOOTH, A. D. & BRITTEN, K. H. V. (1948). *Proc. Roy. Soc. A*, **193**, 305.
 BRAGG, W. L., HOWELLS, E. R. & PERUTZ, M. F. (1952). *Acta Cryst.* **5**, 136.
 COCHRAN, W. (1950). *Acta Cryst.* **3**, 268.
 COCHRAN, W. (1951). *Acta Cryst.* **4**, 81.
 DARWIN, C. G. (1922). *Phil. Mag.* **43**, 800.
 DUNITZ, J. D. (1949). *Acta Cryst.* **2**, 1.
 FISHER, R. A. & YATES, F. (1948). *Statistical Tables*, 3rd ed. London: Oliver and Boyd.
 HAMILTON, W. C. (1955). *Acta Cryst.* **8**, 185.
 HAUPTMAN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136.
 HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210.
 JEFFREY, G. A. (1947). *Proc. Roy. Soc. A*, **188**, 222.
 KAN, G. & COLE, W. F. (1949). *Acta Cryst.* **2**, 38.
 KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131.
 LIPSON, H. & WOOLFSON, M. M. (1952). *Acta Cryst.* **5**, 680.
 PHILLIPS, D. C. (1951). Ph.D. Thesis: University of Wales.
 PHILLIPS, D. C., ROGERS, D. & WILSON, A. J. C. (1950). *Acta Cryst.* **3**, 398.
 ROBERTSON, J. M. (1935). *Proc. Roy. Soc. A*, **150**, 348.
 ROBERTSON, J. M. & WHITE, J. G. (1947). *Proc. Roy. Soc. A*, **190**, 329.
 ROGERS, D. (1950). *Acta Cryst.* **3**, 455.
 ROGERS, D. (1954). *Acta Cryst.* **7**, 628.
 ROGERS, D. (1955). In preparation.
 ROGERS, D. & WILSON, A. J. C. (1953). *Acta Cryst.* **6**, 439.
 TRUTER, M. R. (1950). Private communication.
 WALLWORK, S. C. & STANLEY, K. J. (1954). *Acta Cryst.* **7**, 272.
 WEISS, R. J. (1952). *Proc. Phys. Soc. B*, **65**, 553.
 WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 152.
 WILSON, A. J. C. (1951). *Research, Lond.* **4**, 141.