Acta Cryst. (1964). **17**, 67

The use of high-order moments as statistical criteria for distinguishing between centrosymmetric and non-centrosymmetric structures.

By R. Srinivasan and E. Subramanian, *Department of Physics, University of Madras. Madras 25, India*

(*Received 6 May* 1963)

Introduction

The use of the various statistical criteria for distinguishing between a centrosymmetric and a non-centrosymmetric structure depends on the basic difference in the nature of the intensity distribution for the two cases as was first shown by Wilson (1949). While a comparison of the experimental data with the theoretical distributions such as the $N(z)$ function of Howells, Phillips & Rogers (1950) or the $P(y)$ curves of Ramachandran & Srinivasan (1959) would constitute, in essence, a complete test, other statistical criteria have also been suggested as simple and rapid tests for this purpose. They are, the ratio $\varrho = \langle |F| \rangle^2 / \langle |F|^2 \rangle$, and the variance of the normalized intensity, $v(z) = \langle (z - \langle z \rangle)^2 \rangle$ (Wilson, 1951).

In the course of another investigation it was noticed by the authors that the values of the high-order absolute moments of z are strikingly different for the two types of distribution. One could therefore expect that these would afford us good statistical criteria for distinguishing between the two cases. The values of these high-order moments can be easily obtained from earlier work (*e.g.* Karle & Hauptman, 1953; Rogers & Wilson, 1953). However, they will be deduced independently in the following section in a very simple way making use of a simplified description of the centrosymmetric and non-centrosymmetric distributions. The possibility of this simplification, which does not seem to have been noticed earlier in the literature, arises from the use of a class of distributions known as the gamma distributions. As will be shown below, it enables us to study with advantage the various properties associated with these distributions. In the last section the results of the tests on the high-order moments will be discussed.

A simplified description of the centrosymmetric and the non-centrosymmetric distributions

A gamma distribution is characterized by the probability density function (*e.g.* Weatherburn, 1961)

$$\varphi(x) = e^{-x} x^{l-1} / \Gamma(l) \tag{1}$$

where the range of variable is 0 to ∞, and $\Gamma(l)$ is the well known gamma function

$$\Gamma(l) = \int_0^\infty e^{-t} t^{l-1} dt \, . \tag{2}$$

When a variable x is distributed according to (1) we call it a 'gamma variable' with parameter l, or symbolically describe it as a $\gamma(l)$ variable.

The distribution function for the normalized intensity, z, in the case of a non-centrosymmetric structure, is given by

$$P(z) dz = e^{-z} dz \, . \tag{3}$$

A comparison of (3) with (1) shows that z has a $\gamma(1)$ distribution. On the other hand, for the centrosymmetric case, if we take the variable as $z' = z/2$, we see that, since

$$P(z) dz = (2\pi z)^{-\frac{1}{2}} \exp\,[-z/2] dz \, , \tag{4}$$

$P(z')$ is given by

$$P(z') dz' = z'^{-\frac{1}{2}} (e^{-z'} / \sqrt{\pi}) dz' \tag{5}$$

and therefore z' is a $\gamma(\frac{1}{2})$ variable.

The advantage of this simplification is that several properties connected with the two distributions can be discussed quite simply from the general results applicable to any $\gamma(l)$ variable. Some of these are given below without proof as they are readily available (Weatherburn, 1961).

The expectation value of a $\gamma(l)$ variable is given by l while the rth absolute moment is given by $l(l+1)\ldots(l+r-1)$. The second moment hence equals $l(l+1)$, so that the variance becomes $l(l+1) - l^2 = l$. It can also be shown that the expectation value of the positive square root is $\Gamma(l+\frac{1}{2})/\Gamma(l)$. Another parameter commonly used in statistics is the kurtosis which is the ratio of the fourth moment about the mean to the square of the variance. For a $\gamma(l)$ variable this takes the value $(6/l) + 3$. These are listed in Table 1.

Table 1. *Statistical parameters of a $\gamma(l)$ variable*

Parameter		Value
rth moment:	$\langle x^r \rangle$	$l(l+1)\ldots(l+r-1)$
Variance:	$\langle (x-\bar{x})^2 \rangle$	l
$\langle \lvert \sqrt{x} \rvert \rangle$		$\Gamma(l+\frac{1}{2})/\Gamma(l)$
Kurtosis:	$\langle (x-\bar{x})^4 \rangle / [v(x)]^2$	$(6/l)+3$

We can write down immediately the values of the various parameters for the centrosymmetric and the non-centrosymmetric cases remembering, however, that the variable involved for the former case is $z' = z/2$. The conversion factor can be easily taken into account. The values are listed in Table 2 and may be found to agree with those available in the literature (Karle & Hauptman, 1953).

Table 2. *Statistical parameters connected with the normalized intensity, z, for centrosymmetric and non-centrosymmetric cases derived by Table* 1

Parameter		Non-centrosymmetric	Centrosymmetric
	$r=1$	1	1
	2	2	3
$\langle z^r \rangle$	3	6	15
	4	24	105
	5	120	945
$v(z)$		1	2
$\langle \lvert F \rvert \rangle / \sqrt{\langle I \rangle} = \varrho^{\frac{1}{2}}$		$\sqrt{\pi}/2$	$\sqrt{(2/\pi)}$
$\langle (z-\bar{z})^4 \rangle / [v(z)]^2$		9	15
$\langle (z-\bar{z})^4 \rangle$		9	60

It is also possible to discuss the nature of a $\gamma(l)$ distribution in terms of l. Thus for instance it can be shown that the distribution is asymptotic to the x axis for all values of l, and when $l > 1$ it has a mode at $x = l - 1$. If $l > 2$, it also touches the x axis at the origin while if $1 < l < 2$, it is tangential to the y axis at that point and if $0 < l < 1$, the curve is asymptotic to both the axes and so on.

There are also certain other general results concerning the gamma distributions which are of interest in the present context. For instance, it can be proved quite generally that the sum of two independent gamma variables with parameters l and m is itself a gamma variable with parameter $(l+m)$. That the intensity for a non-centrosymmetric structure is a $\gamma(1)$ variable can, in fact, be deduced from the above result since it is the sum of the squares of the real and imaginary parts each of which (the square) is a $\gamma(\frac{1}{2})$ variable.

Another result which also finds a useful application concerns the distribution of the quotient of two independent gamma variables. Thus (Weatherburn, 1961, p. 158) if x and y are two independent gamma variables with parameters l and m, then the variable $u = x/(x+y)$ has a beta distribution of the first kind (denoted by $\beta_1(l, m)$) defined by

$$P(u) = u^{l-1}(1-u)^{m-1}/B(l, m) \quad (6)$$

while the quotient $v = x/y$ has a beta distribution of the second kind ($\beta_2(l, m)$) defined by

$$P(v) = v^{l-1}/B(l, m)(1+v)^{l+m} \quad (7)$$

where $B(l, m)$ is the well known beta function. The latter result can be used to deduce the distribution of the quotient of the intensities belonging to two independent crystals. The use of this result will be discussed in detail in another paper from this laboratory (Srinivasan, Subramanian & Ramachandran, 1964).

Results and discussion

It may be seen from Table 2 that the difference between the values of the absolute moments is practically insignificant for the low orders while it increases very rapidly as the order increases. Thus for $r = 4$ the centric case has a value of 105, which is more than four times the value for the non-centric case, namely 24. One could therefore expect they would afford us better criteria compared with the earlier ones, namely ϱ and $v(z)$, which involve at the most of the second moment.

However, the one possible limit that exists in practice is that any errors in the values of z would become increasingly important for larger value of r and for this reason it would be advisable not to try too high a value of r. It looks as if $r = 3$ or 4 would be suitable. One could probably try also the fourth moment about the mean for which the difference between the two cases is particularly marked. However, we tried only $\langle z^4 \rangle$ and the results are shown in Table 3. The usual precautions as are applicable to statistical tests were observed while carrying out the calculations.

It can be seen from Table 3 that the test has revealed clearly, in each case, whether the projection is centrosymmetric or not. The only case where the result was ambiguous was for the b projection of the compound P_4S_5.

Table 3. *Observed values of* $\langle z^4 \rangle$

Crystal	Space group	Centric	Non-centric
Theoretical value		105	24
3,3′-Dichloro-4,4′-dihydroxy-diphenylmethane (Whittaker, 1953)	$C2/c$	98 ($h0l$)	—
Ephedrine HCl (Phillips, 1954)	$P2_1$	98 ($h0l$)	21 ($hk0$)
α-Rhamnose (McGeachin & Beevers, 1957)	$P2_1$	144 ($h0l$)	19 ($hk0$)
5-Methoxy-2-nitrophenol (Bartindale *et al.*, 1959)	$P2_12_12_1$	108 ($h0l$)	—
L-Tyrosine HCl (Srinivasan, 1959)	$P2_1$	83 ($h0l$)	21 ($hk0$)
P_4S_5 (van Houten & Wiebenga, 1957)	$P2_1$	54 ($h0l$)	20 ($0kl$)

We also calculated for this particular case some of the other parameters to see their agreement with the theoretical values. The values obtained were: $\langle z^3 \rangle = 11$, $\langle z^2 \rangle = 2{\cdot}7$, $v(z) = 1{\cdot}6$ and $\varrho = 0{\cdot}63$. Thus if we take all these results into account, there is a strong overall indication that the projection is centrosymmetric. The discrepancy observed in the fourth moment is therefore probably due to statistical errors.

Thus it seems that the high-order moments can also be used in practice as statistical criteria. However, it might be mentioned that since the statistical parameters are always subject to errors it would be better to test a number of parameters instead of just one or two. This is likely to eliminate any possible errors.

We wish to thank Prof. G. N. Ramachandran for his keen interest in this work.

References

Bartindale, G. W. R., Crowder, M. M. & Morley, K. A. (1959). *Acta Cryst.* **12**, 111.

Houten, S. van & Wiebenga, E. H. (1957). *Acta Cryst.* **10**, 156.

Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *Acta Cryst.* **3**, 210.

Karle, J. & Hauptman, H. (1953). *Acta Cryst.* **6**, 131.

McGeachin, H. McD. & Beevers, C. A. (1957). *Acta Cryst.* **10**, 227.

Phillips, D. C. (1954). *Acta Cryst.* **7**, 159.

Ramachandran, G. N. & Srinivasan, R. (1959). *Acta Cryst.* **12**, 410.

Rogers, D. & Wilson, A. J. C. (1953). *Acta Cryst.* **6**, 439.

Srinivasan, R. (1959). *Proc. Indian Acad. Sci.* **50**, 19.

Srinivasan, R., Subramanian, E. & Ramachandran, G. N. (1964). To be published.

Weatherburn, C. E. (1961). *A First Course in Mathematical Statistics.* Cambridge University Press.

Whittaker, E. J. W. (1953). *Acta Cryst.* **6**, 714.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318.

Wilson, A. J. C. (1951). *Research, Lond.* **4**, 141.