Acta Cryst. (1955). **8**, 351

The distribution of the mean intensity of a finite group of reflexions. By E. STANLEY,* *Division of Physics, National Research Council, Ottawa, Canada.*

(*Received* 28 *February* 1955)

Wilson (1949) has shown that, ideally, the intensities in any homogeneous group of reflexions (Stanley, 1955) are distributed in one of two ways. In the form given by Howells, Phillips & Rogers (1950) the distribution functions are

$$(1)P(z)dz = \exp(-z)dz \qquad (1)$$

and

$$(\bar{1})P(z)dz = (2\pi z)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}z)dz\,, \qquad (2)$$

where $z = I/\Sigma$ and $\Sigma = \sum f^2$. Both distributions are of the Type III of Pearson (1895) and it follows (Aitken, 1949, p. 129) that the mean value, $\bar{z}$, of a group of n reflexions is distributed as

$$(1)P(\bar{z})_n d\bar{z} = n^n\Gamma(n)^{-1}(\bar{z})^{n-1} \exp(-n\bar{z})d\bar{z} \qquad (3)$$

and

$$(\bar{1})P(\bar{z})_n d\bar{z} = (\tfrac{1}{2}n)^{\frac{1}{2}n}\Gamma(\tfrac{1}{2}n)^{-1}(\bar{z})^{\frac{1}{2}n-1} \exp(-\tfrac{1}{2}n\bar{z})d\bar{z}\,. \qquad (4)$$

The distributions (3) and (4) are identical apart from the factor 2 in the variable n, and

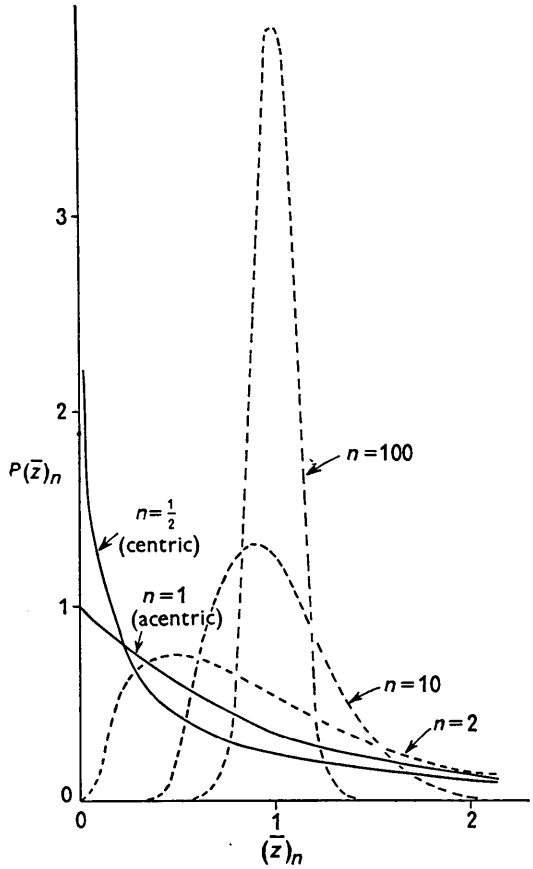

Fig. 1. The distribution of the mean value of n intensities from the acentric distribution.

* National Research Laboratories Postdoctorate Fellow.

$$(1)P(\bar{z})_n = (\bar{1})P(\bar{z})_{2n}\,. \qquad (5)$$

It is convenient to deal only with the distribution (3) since it is simply related to (4) through (5). The distribution is skew but tends asymptotically to the normal distribution as n increases. Fig. 1 shows the distribution graphically for several values of n. The mean value of $\bar{z}$ is

$$\langle\bar{z}\rangle = 1\,. \qquad (6)$$

The most probable value of $\bar{z}$ occurs at

$$(\bar{z})_{\text{max.}} = 1-1/n\,, \qquad (7)$$

and has the value

$$(1)P(\bar{z})_{\text{max.}} = n^n\Gamma(n)^{-1}(1-1/n)^{n-1} \exp(-(n-1))\,. \qquad (8)$$

If n is large, $n! \doteqdot n^n(2\pi n)^{\frac{1}{2}} \exp(-n)$ and this reduces to

$$(1)P(\bar{z})_{\text{max.}} \doteqdot n\{2\pi(n-1)\}^{-\frac{1}{2}}\,, \qquad (9)$$

which is correct to about 1% for $n = 10$.

The variance of $\bar{z}$ is $(1)V(\bar{z})_n = 1/n$, and the standard deviation is

$$(1)\sigma(\bar{z})_n = (1/n)^{\frac{1}{2}}\,. \qquad (10)$$

The fact that $(1)P(\bar{z})_n = (\bar{1})P(\bar{z})_{2n}$ has some interesting consequences. The distribution of the mean values of the reflexions taken in pairs from the centric distribution yields the acentric distribution. This implies that a false distribution would be obtained in any event when non-equivalent reflexions overlap. This could occur in powder photographs of crystals belonging to some space groups and with some twins where two different centric distributions, completely overlapping, would give an acentric distribution. If the overlap is greater than twofold or if the individual distributions are acentric, the resulting distribution will have even lower dispersion.

The $N(z)$ distribution where $N(z) = \int_0^{\bar{z}} P(z)dz$ (Howells *et al.*, 1950) has been used for distinguishing between the distributions (1) and (2). The two curves are somewhat similar in shape. The distribution of $N(\bar{z})$ is

$$(1)N(\bar{z})_n = \int_0^{\bar{z}} P(\bar{z})_n d\bar{z} = \int_0^{\bar{z}} n^n\Gamma(n)^{-1}(\bar{z})^{n-1} \exp(-n\bar{z})d\bar{z}\,. \qquad (11)$$

This is the incomplete gamma function tabulated by Pearson (1934). When $n = 2$

$$(\bar{1})N(\bar{z})_2 = (1)N(\bar{z}) = 1-\exp(-z) \qquad (12)$$

and

$$(1)N(\bar{z})_2 = 4\int_0^{\bar{z}} \bar{z}\exp(-2\bar{z})d\bar{z} = 1-(1+2\bar{z})\exp(-2\bar{z}). \qquad (13)$$

The shape of $(1)N(\bar{z})_2$ is quite different from $(\bar{1})N(\bar{z})_2$, notably in having an opposite initial curvature. These curves are shown in Fig. 2. It is possible that this would provide a more critical test for the distribution type but it would be tedious to apply. For $n > 2$ the family of $N(\bar{z})_n$ curves shown in Fig. 2 all have an initial curvature opposite to that of the $N(z)$ curves.

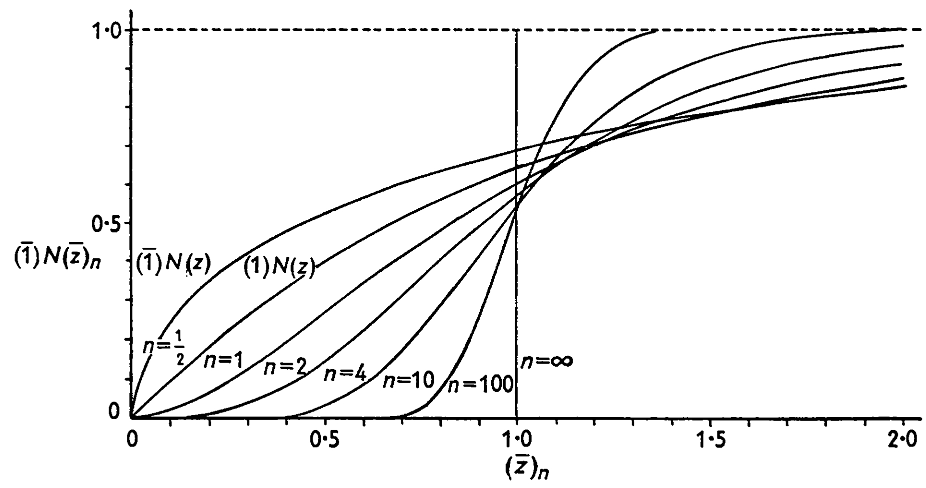

Fig. 2. The cumulative distributions $(1)N(\bar{z})_n$.

The determination of the error in the mean value of the intensities, $\langle I\rangle$, is complicated by the dependence on $\sin\theta$. Provided the range of $\sin\theta$ is not large the standard deviations are

$$(1)\sigma(\langle I_{\text{abs.}}\rangle)_n \doteqdot (1/n)^{\frac{1}{2}}\Sigma$$

and

$$(\bar{1})\sigma(\langle I_{\text{abs.}}\rangle)_n \doteqdot (2/n)^{\frac{1}{2}}\Sigma\,,$$

where $I_{\text{abs.}}$ is the value of the intensity on the absolute scale.

To determine the scaling factor C and the temperature coefficient $2B$, $\ln(\langle I\rangle/\Sigma)$ is plotted against $\sin^2\theta/\lambda^2$ (Wilson, 1942). The standard deviation of the ordinate of each point on the graph is of the order $\sigma(\bar{z})_n$ and the distribution is reasonably symmetrical for $n > 20$. The standard deviation of the extrapolated logarithm of the scaling factor is of the order $\sigma(\bar{z})_n(m-1)^{-\frac{1}{2}}$, where m is the number of equal *unique* groups into which the N reflexions are divided ($N = nm$). The standard deviation of the scaling factor is then of the order $C\sigma(\bar{z})_n(m-1)^{-\frac{1}{2}}$, so that very roughly the standard deviations of the scaling factor for the distributions (1) and (2) are $C(1/N)^{\frac{1}{2}}$ and $C(2/N)^{\frac{1}{2}}$.

The standard error in the temperature coefficient $2B$ will be of the order $\sigma(\bar{z})_n\lambda^2/(m-1)^{\frac{1}{2}}\sin^2\theta_{\text{max.}}$, where $\theta_{\text{max.}}$ is the maximum value of θ used in the graph.

Although this analysis is only approximate it enables some estimate to be made of the reliability of the scaling factor and the temperature factor to be expected from a statistically ideal structure. It may also be of use in indicating whether observed deviations of the individual points on the Wilson graph are within the likely range or due to departures of the structure from the ideal.

The method of Kartha (1953) for determining the absolute scale of intensities suffers from exactly the same type of errors.* Although the equation given by Kartha is exact it involves the summation of the observed intensities to infinity. Any finite summation will have a probable error governed by the distribution type, the way in which the mean intensity varies with $\sin\theta$ and the number of reflexions included in the summation. Since the range of $\sin\theta$ in this case is not small the present analysis is not directly applicable but there seems no reason why the probable error should be less than that in Wilson's method.

I am grateful to Dr D. Rogers for some helpful criticisms.

References

Aitken, A. C. (1949). *Statistical Mathematics*. Edinburgh: Oliver and Boyd.

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

Kartha, G. (1953). *Acta Cryst.* **6**, 817.

Pearson, K. (1895). *Philos. Trans.* **186**, 343.

Pearson, K. (1934). *Tables of the Incomplete Gamma Function*. London: Office of Biometrika, Department of Statistics, University College.

Stanley, E. (1955). *Acta Cryst.* **8**, 356.

Wilson, A. J. C. (1942). *Nature, Lond.* **150**, 152.

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

* Kartha suggests that a temperature coefficient may be obtained from the melting point of the material or from the value of $\sin\theta/\lambda$ at which the mean intensity is reduced to, say, 1/1000 of its maximum value. The first method will give a value of $2B$ which may be very different from the apparent coefficient observed in the X-ray diffraction pattern; the second method is exactly equivalent to, though less precise than, the method of Wilson.