and

$$P_2(\varphi+\pi) \propto \exp[-(\varepsilon_2^2/2E^2)].$$

Similar expressions can be derived for the case shown in Fig. 1(*b*).

Hence, if all phase probabilities other than those at φ and $\varphi+\pi$ are neglected,

$$m=(P_1-P_2)/(P_1+P_2).$$

The quantities P_1 and P_2 can be computed provided there exists a knowledge of the scale factor k. This can be estimated by assuming

$$k\langle f\rangle = 2^{1/2}\langle\Delta\rangle,$$

where $\langle\ \rangle$ signifies mean quantities. Now $\langle f\rangle = J^{1/2}$ if there are J atoms of unit weight in the cell. Hence,

$$k=\langle\Delta\rangle(2/J)^{1/2}.$$

It follows that if, for a particular reflection, $f=0$, then $P_1=P_2$ and $m=0$. Similarly, if $|\Delta|=0$, then $P_1=P_2$ and $m=0$. Indeed, the expression for m is entirely symmetrical between $k|f|$ and $|\Delta|$. The figure of merit is, therefore, small whenever $k|f|$ or $|\Delta|$ are small compared to $\langle\Delta\rangle$.

References

ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* **A31**, 499-506.

ARGOS, P. & ROSSMANN, M. G. (1976). *Acta Cryst.* **B32**, 2975-2979.

ARNOLD, E., ERICKSON, J. W., FOUT, G. S., FRANKENBERGER, E. A., HECHT, H. J., LUO, M., ROSSMANN, M. G. & RUECKERT, R. R. (1984). *J. Mol. Biol.* **177**, 417-430.

BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195-1202.

BRICOGNE, G. (1974). *Acta Cryst.* **A30**, 395-405.

DICKERSON, R. E., KOPKA, M. L., VARNUM, J. C. & WEINZIERL, J. E. (1967). *Acta Cryst.* **23**, 511-522.

GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287-307.

LENTZ, P. J. JR, STRANDBERG, B., UNGE, T., VAARA, I., BORELL, A., FRIDBORG, K. & PETEF, G. (1976). *Acta Cryst.* **B32**, 2979-2983.

RAYMENT, I., JOHNSON, J. E., SUCK, D., AKIMOTO, T. & ROSSMANN, M. G. (1978). *Acta Cryst.* **B34**, 567-578.

ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221-226.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.

ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39-45.

ROSSMANN, M. G., ARNOLD, E., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H. J., JOHNSON, J. E., KAMER, G., LUO, M., MOSSER, A. G., RUECKERT, R. R., SHERRY, B. & VRIEND, G. (1985). *Nature (London)*, **317**, 145-153.

TOLLIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 872-876.

WHITE, J. M. & JOHNSON, J. E. (1980). *Virology*, **101**, 319-324.

Acta Cryst. (1986). A**42**, 334-339

Distributions of Sums and of Ratios of Sums of Intensities

BY A. J. C. WILSON

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

(*Received* 14 *February* 1985; *accepted* 7 *March* 1986)

Abstract

If intensities of reflexion have one or other of the ideal distributions [Wilson (1949). *Acta Cryst.* **2**, 318-320], the sum of n such intensities has a gamma (γ) distribution with parameter n (acentric) or $n/2$ (centric), and the ratio of two such sums has a beta (β) distribution. These distributions are applied to (i) intensities normalized to the ideal average Σ; (ii) intensities normalized to the local average $\langle I\rangle$; (iii) ratios used for scaling. Bias in scaling is discussed, and certain results are obtained for non-ideal distributions. Expressions are obtained for the variance of the traditional reliability index R for both ideal distributions; these have applications in certain methods of structure determination [Rabinovich & Shakked (1984). *Acta Cryst.* **A40**, 195-200].

1. Introduction

1.1. *Notation*

In several crystallographic contexts it is necessary to consider sums like

$$J_n=\sum_{i=1}^{n} G_i, \tag{1}$$

$$K_m=\sum_{i=1}^{m} H_i, \tag{2}$$

and ratios like

$$S_{n,m}=J_n/K_m, \tag{3}$$

where G_i and H_i are the intensities of sets of reflexions. Similar expressions where G_i and H_i are

0108-7673/86/050334-06$01.50

sets of moduli of structure factors also occur, but these are ordinarily less tractable mathematically. An example of the use of a sum is the use of the average intensity,

$$\langle I \rangle = J_n / n, \tag{4}$$

in order to place measured intensities on an approximate absolute scale (Wilson, 1942) before beginning a structure analysis. The ratio $S_{n,m}$ is used in correlating intensities measured under different conditions, such as the same reflexions from different crystals, or the reflexions on different layer lines from the same or different crystals. Such ratios also occur in certain systematic trial-and-error methods of structure determination (see, for example, Rabinovich & Schmidt, 1966; Hirshfeld, 1968; Rabinovich & Shakked, 1984). There are two main cases:

(i) G_i and H_i refer to the *same* reflexion; for example, they might be the observed and calculated quantities for the *hkl* reflexion of a nearly correct structure, or the observed quantities for the *hkl* reflexion measured under different conditions or for different crystals of the same substance; or

(ii) G_i and H_i are *unrelated*, for example, the observed and calculated values for the *hkl* reflexion for a completely wrong trial structure, or values for entirely different reflexions, as in reducing photographic measurements on different layer lines to the same scale.

Aside from the scale factor, in case (i) G_i and H_i will differ chiefly through relatively small statistical fluctuations and uncorrected systematic errors, whereas in case (ii) the differences will be relatively large because of inherent differences in the intensities. Here we are concerned only with the unrelated case (ii).

As has been indicated, the results have applications to the determination of absolute and relative scales (§§ 2.2.1, 2.3, 3.2, 3.3) and to a particular method of structure determination (§§ 4.1.1, 4.1.2). There is little in the crystallographic literature concerning the probability distribution of sums like (1) or ratios like (3); certain results are reviewed by Srinivasan & Parthasarathy (1976, ch. 5), but with a bias toward partially related structures that makes it difficult to apply them to the immediate problem.

1.2. *The acentric and centric distributions*

There is an extensive theory of the probability distribution for individual intensities [for references see, for example, Shmueli & Wilson (1981) and Weiss, Shmueli, Kiefer & Wilson (1985)]. Although these probability distributions can be very complex, the central-limit approximations (Wilson, 1949) apply with sufficient accuracy [particularly if the intensities are normalized with respect to the smoothed local average intensity (§§ 2.3 and 3.3 below) rather than with respect to the sum of the squares of the moduli of the atomic scattering factors (French & Wilson, 1978; Wilson, 1981)] in a sufficient number of cases for distributions of J and S based on them to be of interest. The distributions are (acentric)

$$p(I)\,\mathrm{d}I = \exp(-I/\Sigma)\,\mathrm{d}(I/\Sigma) \tag{5}$$

for non-centrosymmetric crystals, and (centric)

$$P(I)\,\mathrm{d}I = (2\Sigma/\pi I)^{1/2} \exp(-I/2\Sigma)\,\mathrm{d}(I/2\Sigma) \tag{6}$$

for centrosymmetric. Probability distributions for J and S based on (5) and (6) are readily obtained by the use of characteristic functions (Cramér, 1945, ch. 10), but the labour is unnecessary, as the required functions have been extensively studied in statistics. They can be described either by gamma (γ) distributions and beta (β) distributions of the second kind, or by χ^2 distributions and Fisher distributions (Cramér, 1945, ch. 18). Either representation involves some inconvenient factors of 2 or $\frac{1}{2}$; the first will be used here because (i) the nuisance factors appear to be fewer, and (ii) one application involves β distributions of the first kind also, and greater homogeneity of treatment is attained. The main properties of the γ and the β distributions are collected in Table 1.

Table 1. *Some properties of γ and β distributions*

If $x_1, x_2, \ldots, x_n$ are independent γ-distributed variables with parameters $p_1, p_2, \ldots, p_n$, their sum is a γ-distributed variable with $p = p_1 + p_2 + \ldots + p_n$.
If x and y are independent γ-distributed variables with parameters p and q, then the ratio $u = x/y$ has the distribution $\beta_2(u; p, q)$. With the same notation, the ratio $v = x/(x+y)$ has the distribution $\beta_1(v; p, q)$.
Differences and products of γ-distributed variables do not lead to simple results.
For proofs, details and references see Kendall & Stuart (1977).

Name of the distribution and functional form	Mean	Variance
γ distribution with parameter p $\gamma_p(x) = [\Gamma(x)]^{-1} x^{p-1} \exp(-x)$ $0 \le x \le \infty,\ p > 0$	p	p
β distribution of first kind with parameters p and q $\beta_1(x; p, q) = [\Gamma(p+q)/\Gamma(p)\Gamma(q)] x^{p-1} \times (1-x)^{q-1}$ $0 \le x \le 1,\ p, q > 0$	$p/(p+q)$	$pq/(p+q)^2(p+q+1)$
β distribution of second kind with parameters p and q $\beta_2(x; p, q) = [\Gamma(p+q)/\Gamma(p)\Gamma(q)] x^{p-1} \times (1+x)^{-p-q}$ $0 \le x \le \infty;\ p, q > 0$	$p/(q-1)$	$p(p+q-1)/(q-1)(q-2)$

1.3. *Polynomial expansions*

The γ and β distributions could be regarded, if necessary, as the first terms in expansions of orthogonal polynomials. The polynomials are Laguerre for γ distributions (Szegö, 1939, ch. V;

Abramowitz & Stegun, 1964, formula 22.2.12) and Jacobi for β (Szegö, 1939, ch. IV; Abramowitz & Stegun, 1964, formula 22.2.2). The use of the simplest type of Laguerre polynomial has been noted already by Shmueli & Wilson (1981).

2. Applications to the acentric distribution

2.1. *Average of n intensities*

Obviously, from (5) and the definition of the γ distribution in Table 1, the acentric distribution is given by

$$p(I)\,\mathrm{d}I = \gamma_1(I/\Sigma)\,\mathrm{d}(I/\Sigma). \quad (7)$$

The sum of n intensities, J_n, thus has the distribution

$$p(J_n)\,\mathrm{d}J_n = \gamma_n(J_n/\Sigma)\,\mathrm{d}(J_n/\Sigma). \quad (8)$$

The average of n intensities has been represented by $\langle I\rangle$ in (4), but to simplify the following equations it will henceforward be written as $Y = J_n/n$. From (8), Y has the distribution

$$p(Y)\,\mathrm{d}Y = \gamma_n(nY/\Sigma)\,\mathrm{d}(nY/\Sigma). \quad (9)$$

The expected value of Y is thus

$$\mu = n\Sigma/n = \Sigma \quad (10)$$

with variance

$$\sigma^2 = n\Sigma^2/n^2 = \Sigma^2/n \quad (11)$$

- as would be expected.

2.2 *Ratio of two intensity averages*

2.2.1. The ratio of two such means is more interesting. From Table 1, the distribution of

$$u = nY/mZ \quad (12)$$

is

$$p(u)\,\mathrm{d}u = \beta_2(nY/mZ;\, n,\, m)\,\mathrm{d}(nY/mZ), \quad (13)$$

where n is the number of intensities included in the numerator and m is the number in the denominator. The expected value of Y/Z is then

$$\mu = [p/(q-1)](m/n) = m/(m-1) = 1 + m^{-1} + \ldots \quad (14)$$

with variance

$$\sigma^2 = [p(p+q-1)/(q-1)^2(q-2)](m^2/n^2)$$
$$= (n+m-1)m^2/(m-1)^2(m-2)n. \quad (15)$$

One sees that Y/Z is a biased estimate of the scaling factor between the two sets of intensities, and that the bias, of the order of m^{-1}, depends only on the number of intensities averaged in the denominator. This may seem odd at first sight, but it becomes plausible when one remembers that the mean of a quantity is an unbiased estimator of itself, but the reciprocal of a mean is not an unbiased estimator of the mean of a reciprocal. The mean exists only if $m > 1$ and the variance only for $m > 2$.

2.2.2. This bias is readily removed by multiplying Y/Z by $(m-1)/m$. Many methods of estimating scaling factors - perhaps most - also introduce bias (Wilson, 1975; Lomer & Wilson, 1975; Wilson; 1976) that is not so easily removed.

2.2.3. For the same numbers of reflexions, the bias in μ and the variance for the centric distribution (§ 3.1 below) are considerably larger than for the acentric distribution. For both distributions the variance of the scaling factor approaches zero when n and m become large.

2.3. *Intensities normalized to local average*

The distribution of the ratio $v = x/(x+y)$ is chiefly of interest when x relates to a single reflexion and y relates to the remaining $(n-1)$ of a group of n intensities. This corresponds to normalizing intensities to the local average Y instead of to Σ; v is then the normalized intensity. Its distribution is (Table 1)

$$p(v)\,\mathrm{d}v = \beta_1(I/nY;\, 1,\, n-1)\,\mathrm{d}(I/nY), \quad (16)$$

with an expected value of I/Y of

$$\mu = [p/(p+q)]n = 1; \quad (17)$$

there is no bias, as is obvious *a priori*. The variance of I/Y is

$$\sigma^2 = [pq/(p+q+1)(p+q)^2]n^2 = (n-1)/(n+1), \quad (18)$$

which is less than the variance of the intensities normalized to an 'infinite' population by a fraction of the order of $2/n$. Unlike the variance of the scaling factor, the variance of the normalized intensity approaches unity as n becomes large. Also, on taking the limit for n large, the β distribution (16) tends to $\gamma_1(I/Y)\,\mathrm{d}(I/Y)$, the centric distribution with the mean equal to the local average instead of to Σ.

3. Applications to the centric distribution

3.1. *Average of n intensities*

From (6) and Table 1, the centric distribution is

$$p(I)\,\mathrm{d}I = \gamma_{1/2}(I/2\Sigma)\,\mathrm{d}(I/2\Sigma). \quad (19)$$

The centric equivalents of the acentric equations (8)-(10) are thus obtained by replacing n by $n/2$ and Σ by 2Σ, giving

$$p(J_n)\,\mathrm{d}J_n = \gamma_{n/2}(J_n/2\Sigma)\,\mathrm{d}(J_n/2\Sigma), \quad (20)$$

and the average of n intensities, $Y = J_n/n$, has the

distribution

$$p(Y)\,dY = \gamma_{n/2}(nY/2\Sigma)\,d(nY/2\Sigma). \quad (21)$$

The expected value of Y is then

$$\mu = (n/2)(2\Sigma/n) = \Sigma \quad (22)$$

with variance

$$\sigma^2 = (n/2)(2\Sigma/n)^2 = 2\Sigma^2/n. \quad (23)$$

3.2. *Ratio of two intensity averages*

In the centric case, in the expression for the distribution of the ratio of two means Y and Z, p becomes $n/2$ and q becomes $m/2$, so that

$$p(u)\,du = \beta_2(nY/mZ;\ n/2,m/2)\,d(nY/mZ), \quad (24)$$

with the expected value of Y/Z equal to

$$\mu = [p/(q-1)](m/n) = m/(m-2) = 1+2m^{-1}+\ldots \quad (25)$$

and with its variance equal to

$$\sigma^2 = [2n(n+m-2)/(m-2)^2(m-4)](m^2/n^2)$$
$$= 2(n+m-2)m^2/(m-2)^2(m-4)n. \quad (26)$$

The variance is thus large for m small, in fact 'infinite' if the number of terms averaged in the denominator is less than five, but goes to zero for large m and n.

3.3. *Intensities normalized to the local average*

The distribution of intensities normalized to the local average is given by

$$p(v)\,dv = \beta_1[I/nY;\ \tfrac{1}{2},(n-1)/2]\,d(I/nY), \quad (27)$$

with an expected value of I/Y of

$$\mu = [p/(p+q)]n = 1 \quad (28)$$

with variance

$$\sigma^2 = [pq/(p+q+1)(p+q)^2]n^2 = 2(n-1)/(n+2), \quad (29)$$

less than that for an 'infinite' population by a fraction of about $3/n$. The limit of the distribution (27) for n large is the centric distribution $\gamma_{1/2}(I/2Y)\,d(I/2Y)$ with mean Y instead of Σ; the passage to the limit is rather more difficult than for the acentric distribution (§ 2.3).

3.4. *Large intensities*

For intensities approaching the maximum possible in centrosymmetric space groups the modulus F of the structure factor is distributed as a power of $(1-U)$, where

$$U = F/\varphi, \quad (30)$$

$$\varphi = \sum_{i=1}^{N} f_i, \quad (31)$$

and f_i is the atomic scattering factor of the ith of the N atoms in the unit cell (Wilson, 1983; Weiss & Kiefer, 1983; Weiss, Shmueli, Kiefer & Wilson, 1985). The result of Weiss & Kiefer for $P\bar{1}$ is

$$p(U)\,dU = \left[(2\pi)^{N/4}\Gamma(N/4)\prod_{i=1}^{N/2}(2f_i/\varphi)^{1/2}\right]^{-1} \times (1-U)^{N/4-1}\,dU. \quad (32)$$

Aside from the constant factor, this is a β distribution of the first kind, with $p=1$, $q=N/4$. [It is not to be expected that the constant factors will agree, as the distribution (32) represents $p(U)$ only for U approaching unity.]

4. Other topics

4.1. *Variance of R for wrong structures*

4.1.1. The method of structure determination developed by Rabinovich & Shakked (1984) involves evaluation of the traditional R,

$$R = \sum_{i=1}^{n} X_i \Big/ \sum_{i=1}^{n} F_i, \quad (33)$$

for a series of 'random' trial structures; F_i is the observed structure factor and X_i is the difference between the observed and calculated values of the ith structure factor. The expected values of R for wrong structures are (Wilson, 1950) $2-2^{1/2} = 0{\cdot}586$ for F's having the acentric distribution and $8^{1/2}-2 = 0{\cdot}828$ for F's having the centric distribution. The structure factors calculated for the trial structures can be regarded as a random sample from the possible structure factors, and the values of R obtained in practice scatter above and below the theoretical values just quoted, particularly when the number of reflexions that can be used is small. During a conversation in December 1983, Professor Rabinovich asked me how large such statistical fluctuations are likely to be, a question that led to the train of thought of the present paper. It is fairly easy to determine the statistical standard deviation of R if one adopts the 1950 standpoint, although in 1984 certain difficult refinements might be suggested.

4.1.2. The variance of R should be, from (33),

$$\sigma^2(R) = \left[\sum_{i=1}^{n}\sigma^2(X_i)\right] \Big/ \left[\sum_{i=1}^{n} F_i\right]^2 \quad (34)$$

$$= \sigma^2(X)/n\langle F\rangle^2, \quad (35)$$

since all the F's and X's have the same distribution. Thus

$$\sigma^2(X) = \langle X^2\rangle - \langle X\rangle^2 \quad (36)$$

$$= \langle (F_o^2 - 2F_oF_c + F_c^2)\rangle - \langle X\rangle^2 \quad (37)$$

$$= 2\langle F^2\rangle - 2\langle F\rangle^2 - \langle X\rangle^2. \quad (38)$$

From (33) it follows that

$$\langle F\rangle = \langle F\rangle\langle R\rangle, \tag{39}$$

so that

$$\sigma^2(X) = 2\Sigma - \langle F\rangle^2(2+\langle R\rangle^2) \tag{40}$$

and

$$\sigma^2(R) = n^{-1}[2\Sigma/\langle F\rangle^2 - 2 - \langle R\rangle^2]. \tag{41}$$

For acentric distributions

$$\langle F\rangle = \tfrac{1}{2}(\Sigma\pi)^{1/2}, \qquad \langle R\rangle = 2-2^{1/2}, \tag{42}$$

so that

$$\sigma^2(R) = n^{-1}[(8/\pi)+(32)^{1/2}-8] = 0{\cdot}203/n \tag{43}$$

and for centric distributions

$$\langle F\rangle = (2\Sigma/\pi)^{1/2}, \qquad \langle R\rangle = 8^{1/2}-2, \tag{44}$$

so that

$$\sigma^2(R) = n^{-1}[\pi+(128)^{1/2}-14] = 0{\cdot}455/n. \tag{45}$$

If n is not large the corresponding standard deviations can be quite big; for example, if $n=25$ the value of the acentric R is 0·59 with standard deviation 0·09 and that of the centric R is 0·83 with standard deviation 0·13.

4.2. *Some results for non-ideal distributions*

4.2.1. *The median of a ratio.* From the symmetry relation for the incomplete β function (Abramowitz & Stegun, 1964, formula 6.6.3) it follows immediately that the median of the β_1 distribution is at $x=\frac{1}{2}$, whatever the values of p and q. The relation between the two β distributions (Table 1) then implies that the median of the β_2 distribution is at $x=1$. This distribution results from the ratio of two variables with γ distributions, and one is therefore led to ask if a similar result holds for ratios of non-negative variables with other distributions. The following argument appears to show that it does hold if the numerator and denominator have the same distribution, but not necessarily otherwise. The distribution function of a ratio $u=x/y$ is given by (Kendall & Stuart, 1977, p. 282)

$$p(u)\,\mathrm{d}u = \int_0^\infty q_1(uy)q_2(y)\,y\,\mathrm{d}y\,\mathrm{d}u, \tag{46}$$

where the q's are the probability density functions of x and y. If the q's are different there is no obvious way forward, but progress can be made if $q(.)$ is the same for both x and y. The median is the value of m that satisfies the equation

$$\tfrac{1}{2} = \int_0^m \int_0^\infty q(uy)q(y)y\,\mathrm{d}y\,\mathrm{d}u. \tag{47}$$

Substitution of $t=uy$ and integration first with respect to u (t) gives

$$\tfrac{1}{2} = \int_0^\infty \int_0^{my} q(t)\,\mathrm{d}t q(y)\,\mathrm{d}y \tag{48}$$

$$= \int_0^\infty Q(my)q(y)\,\mathrm{d}y, \tag{49}$$

where $Q(.)$ is the cumulative distribution function corresponding to $q(.)$. In general there is no simple integral, but it is obvious that $m=1$ is a solution, whatever the form of $q(.)$. With $m=1$, (48) becomes

$$\tfrac{1}{2} = \int_0^\infty Q(y)\,\mathrm{d}[Q(y)] = 2^{-1}[Q(y)]^2]_0^\infty = \tfrac{1}{2}, \tag{50}$$

an identity. The median is therefore at $u=1$, whatever the form of $q(.)$. We thus find that when $n=m$, Y/Z is an unbiased estimate of the median of the distribution of the scale factor, even though it is not an unbiased estimate of the mean.

4.2.2. *Bias of the scaling ratio.* With Y and Z defined by [(1) and (2)]

$$Y = J_n/n, \qquad Z = K_m/m, \tag{51}$$

the scaling ratio u is

$$u = Y/Z \tag{52}$$

in any particular case, and has the expected value

$$\langle u\rangle = \langle Y/Z\rangle = \langle Y\rangle\langle Z^{-1}\rangle. \tag{53}$$

The expected value of Y is the mean *population* intensity, say M, and is not to be confused with Y, which is the mean of a particular *sample* of n intensities. If the probability distribution of Z is $q(Z)$, the required value of Z^{-1} is

$$\langle Z^{-1}\rangle = \int Z^{-1}q(Z)\,\mathrm{d}Z. \tag{54}$$

This is not equal to M^{-1}; fluctuations to small values increase Z^{-1} by more than fluctuations to high values decrease it, so that $\langle Z^{-1}\rangle$ is always greater than M^{-1}, and is in fact infinite for small values of m (1 and 2 for the ideal acentric and centric distributions, as was seen in §§ 2.2.1 and 3.2, but 4 or more for many non-ideal distributions).

The effect just described depends on the spread of the values of Z about its expected value M, a spread which is measured by the variance of Z. The ratio of the discrepancies for the ideal distributions is the same as the ratios of their variances, suggesting that in general the bias will be of the order of

$$1+\sigma^2(I)/m\langle I\rangle^2, \tag{55}$$

whatever the intensity distribution. This conjecture is made plausible by the following argument. The variance of Z is, of course, equal to the variance of a single intensity I divided by m, so that for larger m the distribution becomes more and more concentrated

about $Z = M$. One is therefore led to expand Z^{-1} in powers of $(Z-M)/M$, giving

$$Z^{-1} = M^{-1}[1-(Z-M)/M+(Z-M)^2/M^2-\ldots + (-)^k(Z-M)^k/M^k+\ldots] \quad (56)$$

$$= M^{-1}-0+\sigma^2(Z)/M^3+\ldots + (-)^k\mu_k(Z)M^{k+1}+\ldots, \quad (57)$$

$$\langle Y/Z\rangle = 1+\sigma^2(I)/m\langle I\rangle^2-\ldots + (-)^k\mu_k(Z)/\langle I\rangle^k+\ldots; \quad (58)$$

the first two terms agree with the conjecture (55).

The moments μ_k for k odd are small. Regarded as functions of k, those for k even increase with k, and the series may not converge. However, regarded as functions of m, they are of the order of m^{-k-1}, and the terms of the series (58) decrease with increasing m. The series thus appears to be asymptotic; for fixed m the terms at first decrease with increasing k, but ultimately increase. Such series give a good approximation if only terms in the decreasing region are used.

Values of $\sigma^2(I)$ as a function of space group and atomic composition have been given by Wilson (1950, 1978), Foster & Hargreaves (1963), Shmueli & Kaldor (1981) and Shmueli & Wilson (1981). Even if there is no atomic heterogeneity the effect of symmetry can be appreciable. Wilson (1978, § 1.5) considered the hypothetical case of an element crystallizing in three allotropic forms, each with 24 atoms in the unit cell, and having respectively space groups $P1$, $P2/m$, and $P6/mmm$. The coefficients of m^{-1} in (55) would be 0·958, 1·875 and 3·125. The first two are close to the ideal values 1 and 2, but the third is markedly higher.

4.3. *The use of normal approximations*

Since J_n and K_m [(1) and (2)] are sums of identically distributed variables conforming to the conditions of the central-limit theorem, it is tempting to approximate their distributions by normal distributions with the correct mean and variance. This would be reasonably satisfactory for the distributions of J_n and K_m themselves for quite small values of n and m, but unsatisfactory for the distribution of their ratio for any values of n and m, even large. The ratio of two variables with normal distributions is notorious for its rather indeterminate mean and 'infinite' variance, resulting from the 'tail' of the denominator distribution extending through zero to negative values. The leading term of the ratio distribution is given by Kendall & Stuart (1977, p. 288).

References

ABRAMOWITZ, M. & STEGUN, I. A. (1964). *Handbook of Mathematical Functions.* Washington, DC: US Government Printing Office.

CRAMÉR, H. (1945). *Mathematical Theory of Statistics.* Uppsala: Almqvist & Wiksells.

FOSTER, F. & HARGREAVES, A. (1963). *Acta Cryst.* **16**, 1133-1139.

FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* A**34**, 517-525.

HIRSHFELD, F. L. (1968). *Acta Cryst.* A**24**, 301-311.

KENDALL, M. & STUART, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. London; Griffin.

LOMER, T. R. & WILSON, A. J. C. (1975). *Acta Cryst.* B**31**, 646-648.

RABINOVICH, D. & SCHMIDT, G. M. J. (1966). *Nature (London)*, **211**, 1391-1393.

RABINOVICH, D. & SHAKKED, Z. (1984). *Acta Cryst.* A**40**, 195-200.

SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* A**37**, 76-80.

SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 342-353.

SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon.

SZEGŐ, G. (1939). *Orthogonal Polynomials.* New York: American Mathematical Society.

WEISS, G. H. & KIEFER, J. E. (1983). *J. Phys. A*, **16**, 489-495.

WEISS, G. H., SHMUELI, U., KIEFER, J. E. & WILSON, A. J. C. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 23-42. Guilderland, NY: Adenine Press.

WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151, 152.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-320.

WILSON, A. J. C. (1950). *Acta Cryst.* **3**, 397-398.

WILSON, A. J. C. (1975). In *Anomalous Scattering*, edited by S. RAMASESHAN & S. C. ABRAHAMS, pp. 325-332. Copenhagen: Munksgaard.

WILSON, A. J. C. (1976). *Acta Cryst.* A**32**, 994-996.

WILSON, A. J. C. (1978). *Acta Cryst.* A**34**, 986-994.

WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 808-810.

WILSON, A. J. C. (1983). *Acta Cryst.* A**39**, 26-28.