Acta Cryst. (1965). 19, 1003

Probability Distribution Connected with Structure Amplitudes of Two Related Crystals. IV. The Distribution of the Normalized Difference*

BY R. SRINIVASAN AND G. N. RAMACHANDRAN

Centre of Advanced Study in Physics, University of Madras, Madras 25, India

(*Received* 9 *February* 1965 *and in revised form* 1 *May* 1965)

The theory developed in the earlier parts of this series concerning the probability distribution of structure amplitudes is particularly applicable to the case of the 'observed' and 'calculated' structure factors, during the initial stages of a crystal structure analysis when only a part of the structure is known. In such a case the proper type of difference variable to consider is the *normalized* difference variable $\delta=(|F_N|/\sigma_N-|F_P|/\sigma_P)$ (where σ_N^2 and σ_P^2 are the mean square values of $|F_N|$ and $|F_P|$respectively). The distribution of δ is derived, as in previous parts, for the related and unrelated cases and is found to have all the expected symmetry properties like the normalized product and quotient variables considered in parts II and III respectively.

Based on the distribution $P(\delta)$, a new reliability index R_1, termed the normalized reliability index, is suggested for use in crystal-structure refinement when only a part of the structure is known. R_1, defined $R_1=\Sigma\ ||F_N|-|F_P|/\sigma_1|/\Sigma\ |F_N|$, tends to the conventional reliability index R when P tends to N. R_1 has many desirable properties which makes it superior to R when P is less than N.

1. Introduction

The basic problem considered in parts I to III of this series (part I: Ramachandran, Srinivasan & Sarma, 1963; part II: Srinivasan, Sarma & Ramachandran, 1963*a*; part III: Srinivasan, Subramanian & Ramachandran, 1964; see also Ramachandran & Srinivasan, 1963 and Srinivasan, Sarma & Ramachandran, 1963*b*) was to work out the probability distributions connected with structure amplitudes $|F_N|$ and $|F_P|$ of two crystals containing N and P atoms respectively. Two limiting situations were considered, one in which P out of the N atoms in the first crystal occupied positions identical with those of the corresponding P atoms of the second crystal (termed the related case) and the other in which the P atoms and the N atoms occupied completely different positions (termed the unrelated case). The results thus obtained formed the basis for suggesting a number of statistical criteria for use in testing the 'relatedness' or 'isomorphism' between two crystals.

The purpose of this paper is mainly twofold. First, it may be mentioned that there is another but equivalent interpretation of the mathematical results obtained earlier which is particularly significant. It arises from considering P and N not as corresponding to two different crystals but as a part P and the whole N of the same crystal, which would correspond to the 'related case' considered earlier. (In fact, this was how the whole problem came to be considered in part I, although the emphasis in the earlier parts has been mainly on applying the results to a pair of crystals.) Thus, one could take $|F_P|$ to represent the calculated structure amplitude (say $|F_c|$) corresponding to a part consisting of P atoms, which have been located initially during a structure analysis, and $|F_N|$ to correspond to $|F_o|$, the observed structure amplitude of the whole crystal (the errors of observation are assumed to be negligible). Thus, it becomes possible to apply the results directly for the comparison of $|F_o|$ and $|F_c|$ during the different stages of analysis of a crystal structure, when more and more atoms are located. It may be perhaps emphasized that the application of the theory to this case has a greater validity than to the case of a pair of crystals, since the latter very rarely possess identical cell dimensions, as is implicitly required by the theory. The 'unrelated case' clearly would correspond to the case in which the assumed positions of the P atoms are entirely different from their true positions.

Secondly, the theoretical treatment of the problem and the expressions for the distributions of different types of variables turn out to be quite elegant when the two variables F_N and F_P are taken individually in their normalized forms, namely $y_N=|F_N|/\sigma_N$ and $y_P=|F_P|/\sigma_P$. Thus, the quotient variable, $v=y_N/y_P$ considered in part III, is in the correctly normalized form. It showed highly symmetric properties. For instance, the distribution of v and its reciprocal $1/v$ were identical in form and also for any given value of σ_1^2, the integral $\int_1^\infty P(v)dv$ always had the value $\frac{1}{2}$. These results were not true for the direct quotient of the two structure amplitudes, $|F_N|/|F_P|$. So also, the product variable $(Z=y_Ny_P)$ was used in the normalized form (part II), although in this case no symmetry properties could be demonstrated as in the case of the normalized quotient variable. Also, both in the case of $P(v)$ and $P(Z)$, it was found that the expression for the unrelated case

* Contribution 165 from the Centre of Advanced Study in Physics, University of Madras.

could be obtained from the related case by substituting $\sigma_1^2=0$.

It may be mentioned that the difference variable considered in part I was defined as $w=(|F_N|-|F_P|)/\sigma_N$ and is obviously not in its properly normalized form. In view of the above it would be of particular interest to consider the normalized difference* variable

$$\delta=y_N-y_P\,. \tag{1}$$

This paper is concerned with working out the distributions $P(\delta)$. As expected, $P(\delta)$ is found to have all the symmetry properties of $P(v)$ and $P(Z)$. So also, as will be shown in part V, in which the entire problem is considered in its general aspect, namely when the P atoms have also finite errors in their coordinates, very elegant expressions are obtained for the distribution functions with the normalized difference variable δ which is introduced here.

One important result that follows from considering the distribution $P(\delta)$ is the possibility of a new reliability index (termed the normalized reliability index) for use in crystal structure refinements when only a part of the atoms is known for calculating the structure factors. The new index is found to have a number of desirable properties which make it superior to the conventional one.

2. The probability distribution of δ

Non-centrosymmetric case

We require the distribution of $\delta=y_N-y_P$. Consider first the non-centrosymmetric case. We first obtain the conditional distribution $P_2(y_N;\,y_P)$ for the normalized variable y_N. This is easily done by applying the appropriate transformations in the expression for $P_2(|F_N|;\,|F_P|)$ available in part I (expression 4). Thus,

$$P_2(y_N;\,y_P) =(2y_N/\sigma_2^2)I_0[2\sigma_1 y_N y_P/\sigma_2^2]\exp\{-[(y_n^2+\sigma_1^2 y_P^2/\sigma_2^2)]\} \tag{2}$$

and $P_1(y_P)$ is given by

$$P_1(y_P)=2y_P\exp(-y_P^2)\,. \tag{2a}$$

Thus

$$P(\delta)=\int_{0\text{ or }|\delta|}^{\infty} P_2(y_P+\delta;\,y_P)\,P_1(y_P)\,dy_P\,. \tag{3}$$

Substituting the expressions for $P_2(y_N;y_P)$ and $P_1(y_P)$ from (2) and (2a) in (3) we obtain $P(\delta)$ in the form

$$P(\delta)=2\exp(-\delta^2/\sigma_2^2)\int_{0\text{ or }|\delta|}^{\infty} x\exp(-x)I_0(\sigma_1 x)\,dy_P \tag{4}$$

where

$$x=2y_P(y_P+\delta)/\sigma_2^2 \tag{4a}$$

and the lower limits 0 and $|\delta|$ apply respectively to positive and negative values of δ. That the function $P(\delta)$ above is symmetric about $\delta=0$ can be readily seen as follows. When δ is negative, we have $\delta=-|\delta|$ where $|\delta|$ is positive. Changing the variable of integration to $y_P'=y_P-|\delta|$ we see that $x=2y_P(y_P+\delta)/\sigma_2^2=2(y_P'+|\delta|)y_P'/\sigma_2^2$, which is identical in form with 4(a) for δ positive, and the limits are now from 0 to ∞.

Similarly, for the unrelated case, the distribution $P(\delta)$ may be derived from the corresponding formulae in part I (details not given here) and it takes the form

$$P(\delta)=2\exp(-\delta^2)\int_{0\text{ or }|\delta|}^{\infty} 2y_P(y_P+\delta)\times\exp\{-[2y_P(y_P+\delta)]\}\,dy_P\,. \tag{5}$$

It is readily verified that this expression is obtained from that of the related case (4) above, by substituting $\sigma_1^2=0$ ($\sigma_2^2=1$) in the right hand side of (4).

Fig. 1 shows a family of curves of $P(\delta)$ for different values of σ_1^2.

Centrosymmetric case

From expression (11) of part I we obtain

$$P_2(y_N;\,y_P)=\frac{1}{\sqrt{2\pi\sigma_2^2}}[\exp\{-(y_N+\sigma_1 y_P)^2/2\sigma_2^2\}+\exp\{-(y_N-\sigma_1 y_P)^2/2\sigma_2^2\}]\,. \tag{6}$$

We have

$$P_1(y_P)=\sqrt{\frac{2}{\pi}}\exp(-y_P^2/2)\,. \tag{6a}$$

Hence, substituting in (3) for $P_2(y_N;\,y_P)$ and $P_1(y_P)$ from (6) and (6a), we obtain

$$P(\delta)=\frac{1}{\pi\sigma_2}\int_{0\text{ or }|\delta|}^{\infty}\exp\{-y_P^2/2\}\times\left[\exp\left\{-\frac{[\delta+y_P(1+\sigma_1)]^2}{2\sigma_2^2}\right\}+\exp\left\{-\frac{[\delta+y_P(1-\sigma_1)]^2}{2\sigma_2^2}\right\}\right]dy_P \tag{7}$$

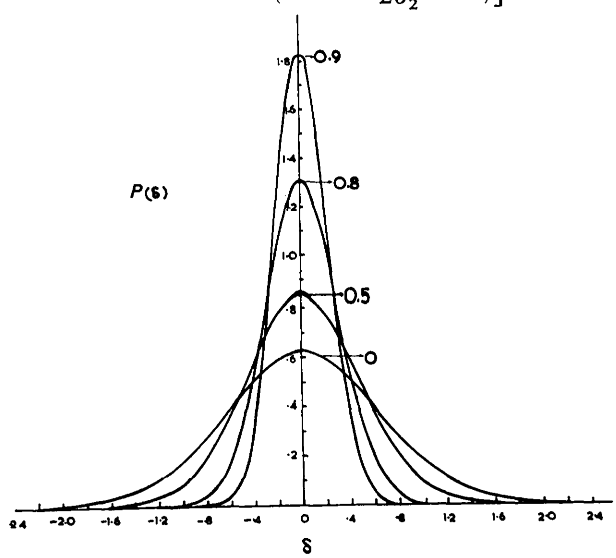

Fig. 1. Probability distribution function $P(\delta)$ for the non-centrosymmetric case, corresponding to $\sigma_1^2=0$, 0·5, 0·8 and 0·9.

---

* Hereafter, we shall refer to w as the simple difference to distinguish it from the normalized difference δ defined here.

where again, as before, the lower limit is 0 or $|\delta|$ according as δ is positive or negative.

The integral may be expressed in terms of error functions. We shall first work out the result for positive δ, so that the range of integration is from 0 to ∞. Using the abbreviations

$$a=\delta/\sigma_2\sqrt{2},\quad k_1=(1+\sigma_1)/\sigma_2\sqrt{2},\quad k_2=(1-\sigma_1)/\sigma_2\sqrt{2} \quad (8)$$

the required integral is

$$\frac{\exp(-a^2)}{\pi\sigma_2}\int_0^\infty [\exp\{-[y_P^2(k_1^2+\tfrac{1}{2})+2ak_1y_P]\} + \exp\{-[y_P^2(k_2^2+\tfrac{1}{2})+2ak_2y_P]\}]dy_P\,. \quad (9)$$

The two terms under the integral sign are similar. Taking the first one

$$\int_0^\infty \exp\{-[y_P^2+(k_1^2+\tfrac{1}{2})+2ak_1y_P]\}\,dy_P \quad (10)$$

$$=\exp\left\{-\frac{a^2k_1^2}{(k_1^2+\frac{1}{2})}\right\}\int_0^\infty \exp\{-[y_P\sqrt{(k_1^2+\tfrac{1}{2})} + ak_1/\sqrt{k_1^2+\tfrac{1}{2}}]^2\}dy_P \quad (11)$$

$$=\frac{\exp\left\{-\frac{a^2k_1^2}{(k_1^2+\frac{1}{2})}\right\}}{\sqrt{(k_1^2+\frac{1}{2})}}\int_{ak_1/\sqrt{k_1^2+\frac{1}{2}}}^\infty \exp(-x^2)dx \quad (12)$$

where

$$x=y_P\sqrt{(k_1^2+\tfrac{1}{2})}+ak_1/\sqrt{k_1^2+\tfrac{1}{2}}\,. \quad (13)$$

Using the result

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^\infty \exp(-t^2)dt\,, \quad (14)$$

(12) reduces to

$$\frac{\sqrt{\pi}}{2}\exp\left\{\frac{-a^2k_1^2}{(k_1^2+\frac{1}{2})}\right\}\left[1-\frac{\operatorname{erf} ak_1}{\sqrt{k_1^2+\frac{1}{2}}}\right]. \quad (15)$$

A similar expression is obtained for the second term under the integral sign in (9), with k_2 replacing k_1 in the above.

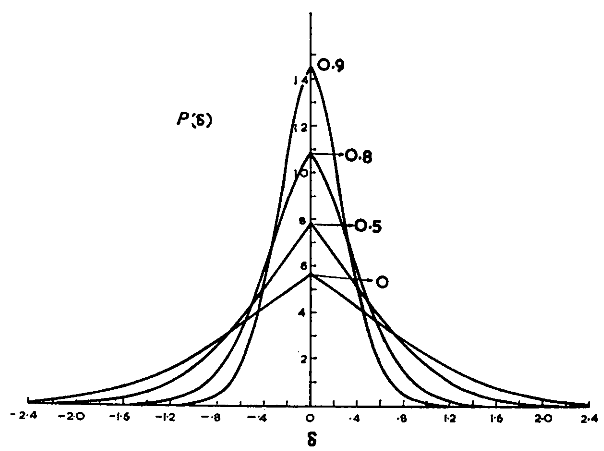

Fig. 2. Probability distribution function $P(\delta)$ for the centrosymmetric case corresponding to $\sigma_1^2=0$, 0·5, 0·8 and 0·9.

After substituting for k_1, k_2, a *etc.* from (8) and simplifying, we obtain finally

$$P(\delta)=\frac{1}{2\sqrt{\pi}}\left\{\frac{\exp(-\delta^2)/4(1+\sigma_1)}{\sqrt{(1+\sigma_1)}}\left(1-\operatorname{erf}\frac{\delta}{2\sqrt{1-\sigma_1}}\right) + \frac{\exp(-\delta^2)/4(1-\sigma_1)}{\sqrt{1-\sigma_1}}\left(1-\operatorname{erf}\frac{\delta}{2\sqrt{(1+\sigma_1)}}\right)\right\}. \quad (16)$$

It may easily be verified by going through the above process that, when δ is negative, the same expression as (17) is obtained with the only change that δ is replaced by $|\delta|$. The function $P(\delta)$ is thus symmetric about $\delta=0$, a result anticipated earlier and shown already for the non-centrosymmetric case.

The expression for the unrelated case may be obtained from the corresponding expression (21) of part I and takes the form

$$P(\delta)=\frac{1}{\sqrt{\pi}}\exp(-\delta^2/4)\left(1-\operatorname{erf}\frac{\delta}{2}\right). \quad (17)$$

Just as in the non-centrosymmetric case, it may be readily verified that the right hand side of (16) for the related case reduces to that of (17) on putting $\sigma_1=0$ ($\sigma_2=1$).

Fig. 2 shows a family of curves of $P(\delta)$ for different values of σ_1^2.

Properties of the $P(\delta)$ distribution

The following observations regarding the properties of the function $P(\delta)$ would be pertinent, particularly in relation to some of the results obtained in earlier parts.

It may be seen from Figs. 1 and 2 that all the curves of $P(\delta)$ are symmetric about $\delta=0$. It follows therefore that

$$\int_{-\infty}^0 P(\delta)\,d\delta=\int_\infty^0 P(\delta)\,d\delta=\tfrac{1}{2}\,. \quad (18)$$

This may be compared with the symmetry property of the quotient variable v about $v=1$, namely

$$\int_0^1 P(v)\,dv=\int_1^\infty P(v)dv=\tfrac{1}{2}\,. \quad (19)$$

Corresponding to the result that the distributions of v and $1/v$ are identical in form, the distribution of $(y_N-y_P)=\delta$ and $(y_P-y_N)=-\delta=\delta'$(say) are the same. The proof of the latter is obvious and is not given here.

The result (18) affords us a very simple method of scaling the observed structure factors. Thus, it is only necessary to adjust the relative scale of $|F_N|$ and $|F_P|$ so that the number of reflexions with δ positive should be equal to the number of reflexions with δ negative (see §3, for a simple method of achieving this practically). This may be compared with the P_+ test discussed in part I for the same purpose, in which the value of P_+ varied with σ_1. This is a consequence of the fact that the simple difference variable is not in its properly normalized form. In fact, this is also reflected in the other properties associated with $P(w)$. For instance, the curves of $P(w)$ are asymmetric about $w=0$ and also a family of curves exists for the unrelated case for different values of σ_1 and so on.

3. The normalized reliability index

Let us now define a reliability index (R_1) between the normalized variables y_N and y_P defined by

$$R_1 = \frac{\Sigma\,|y_N - y_P|}{\Sigma\, y_N} \equiv \frac{\Sigma\,||F_N|/\sigma_N - |F_P|/\sigma_P|}{\Sigma\,|F_N|/\sigma_N}. \quad (19a)$$

From its definition, it is readily seen that

$$R_1 = \langle|\delta|\rangle / \langle y_N \rangle \quad (19b)$$

so that theoretical expressions for this can be readily worked out in terms of the available distributions $P(\delta)$ and $P(y_N)$. In fact, the values of $\langle y_N \rangle$ are well known (Wilson, 1949). It is $\sqrt{\pi}/2$ for a non-centrosymmetric crystal and $\sqrt{2/\pi}$ for a centrosymmetric crystal. Since the distributions $P(\delta)$ are symmetric about $\delta=0$ we have

$$\langle|\delta|\rangle = 2\int_0^\infty \delta P(\delta) d\delta\,. \quad (20)$$

For the non-centrosymmetric case, this can only be obtained by numerical integration using equation (4). However, the expression for $\langle|\delta|\rangle$ takes an elegant form for the centrosymmetric case. Thus, substituting from (16), we have

$$\langle|\delta|\rangle = \frac{1}{\sqrt{\pi}}\left\{\int_0^\infty \frac{\delta}{\sqrt{(1+\sigma_1)}} \exp\left[-\delta^2/4(1+\sigma_1)\right] \operatorname{erfc}\frac{\delta}{2}\sqrt{1-\sigma_1}\,d\delta + \int_0^\infty \frac{\delta}{\sqrt{1-\sigma_1}} \exp\left[-\delta^2/4(1-\sigma_1)\right] \operatorname{erfc}\frac{\delta}{2}\sqrt{1+\sigma_1}\,d\delta\right\} \quad (21)$$

where

$$\operatorname{erfc} x = 1 - \operatorname{erf}(x)\,. \quad (22)$$

The two integrals in (21) are similar. Taking the first one and substituting $\delta^2/4\sqrt{1+\sigma_1} = x$, it reduces to

$$2\sqrt{1+\sigma_1}\int_0^\infty \exp(-x) \operatorname{erfc}\left(\frac{x^{\frac{1}{2}}\sqrt{1+\sigma_1}}{\sqrt{1-\sigma_1}}\right) dx\,. \quad (23)$$

Using the result (see p.307, vol. II of Erdelyi, 1954)

$$\int_0^\infty \exp(\beta x) \operatorname{erfc}(a^{\frac{1}{2}}x^{\frac{1}{2}}) dx = \frac{1}{\beta}\left[\frac{a^{\frac{1}{2}}}{(a-\beta)^{\frac{1}{2}}} - 1\right], \quad (24)$$

(23) reduces to

$$2\sqrt{1+\sigma_1} - \sqrt{2(1+\sigma_1)}\,. \quad (25)$$

Similarly, the second integral in (21) reduces to

$$2\sqrt{1-\sigma_1} - \sqrt{2(1-\sigma_1)}\,. \quad (26)$$

Using (25) and (26) in (21) we have

$$\langle|\delta|\rangle = \frac{2}{\sqrt{\pi}}\left\{\sqrt{1+\sigma_1} + \sqrt{1-\sigma_1} - \sqrt{2}\right\}. \quad (27)$$

Since $\langle y_N \rangle = \sqrt{2/\pi}$ for the present case, we get

$$R_1 = \langle|\delta|\rangle/\langle y_N\rangle = \sqrt{2(1+\sigma_1)} + \sqrt{2(1-\sigma_1)} - 2\,. \quad (28)$$

The value of R_1 as a function of σ_1 is shown in Fig. 3 for the centrosymmetric and non-centrosymmetric cases.

It is interesting to note that the quantity R_1 is closely related to the usual R value which is defined between the observed and calculated structure amplitudes, *viz.*

$$R = \Sigma||F_N| - |F_P||/\Sigma|F_N|\,. \quad (29)$$

Thus, multiplying both the numerator and denominator of the last expression in (19) by σ_N we have

$$R_1 = \Sigma||F_N| - |F_P|\sigma_N/\sigma_P|/\Sigma|F_N| = \Sigma||F_N| - |F_P|/\sigma_1|/\Sigma|F_N|\,. \quad (30)$$

Thus the index R_1 makes use of the differences between the structure amplitude $|F_N|$ of the whole crystal and $|F_P|$ of a part of it, with the only difference that the latter is scaled up by a factor $1/\sigma_1$. This scaling up is physically understandable and in a sense justified since the mean square value of $|F_P|$ is less than that of $|F_N|$. Thus, the scaling up ensures the equality of these quantities because $\langle|F_N|^2\rangle = \langle|F_P|^2/\sigma_1^2\rangle$. Thus, the effect of this is that not only are the structure amplitudes to be compared properly normalized, but the differences between them are also made symmetrical. It may be seen that as P tends to N, $\sigma_1 \to 1$, and R_1 becomes identical with the conventional reliability index R.

The above result also has one important practical significance. For the normalization is equivalent to making the average of the 'observed intensities' equal

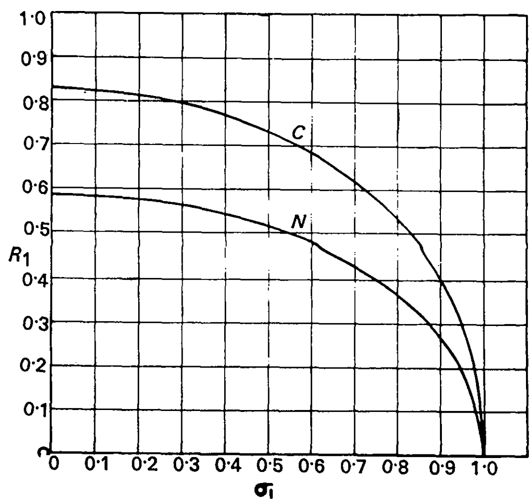

Fig. 3. The normalized reliability index R_1 as a function of σ_1 for the centrosymmetric (C) and non-centrosymmetric (N) cases.

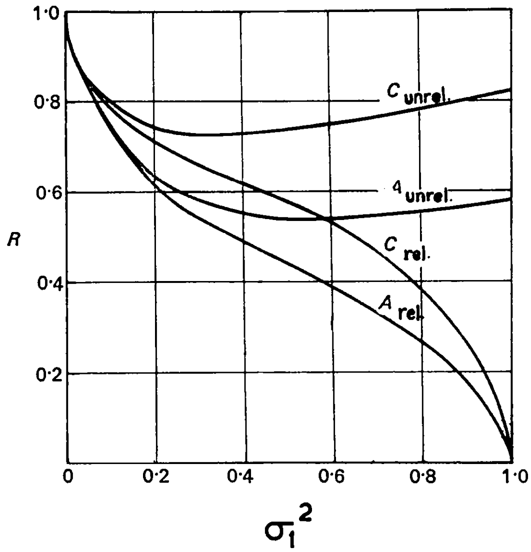

Fig. 4. The reliability index R as a function of σ_1^2 for the related and unrelated cases. A and C denote acentric (non-centrosymmetric) and centrosymmetric cases respectively.

to that of the 'calculated ones' and this can be achieved even without a prior knowledge of the correct scale factor for $|F_o|$ or of even the value of σ_1. What is required is only that the scale factor between the two is adjusted by trial until the above equality is achieved. In this respect R_1 has a distinct advantage over R, which necessarily requires the knowledge of the correct scale factor.

It is of interest to compare further the behaviour of R_1 and R. For convenience, R as a function of σ_1 is shown in Fig. 4. It will be noticed that for the unrelated case the value of R varies with σ_1 while R_1 has a unique value, independent of σ_1. In fact, there is only a single curve for R_1 as a function of σ_1 for the related case, and the unrelated case corresponds to $\sigma_1 = 0$. This result follows the corresponding result for the distributions, and the significance of this will be better appreciated in the light of the results to be discussed in part V, where the errors in atomic coordinates are discussed. It is also noteworthy that these values of R_1 for the centrosymmetric and non-centrosymmetric cases (0·828 and 0·586 respectively) are nothing but those corresponding to completely random structures deduced by Wilson (1950) for the case when all the atoms are used in structure factor calculation.

On the other hand, the value of R for the unrelated case corresponds to that of Wilson, for random structures, only in the limit $P = N$, *i.e.* $\sigma_1 = 1$. For lower values of σ_1 R shows a steady increase until it reaches the value unity when $\sigma_1 = 0$. This happens both for the centrosymmetric and non-centrosymmetric cases. It arises from the fact that when $\sigma_1 = 0$ the $|F_P|$'s are all zero and both the numerator and denominator in (29) become equal.

For the related case, however, R starts with a value unity for $\sigma_1 = 0$, and decreases steadily until it becomes zero when $\sigma_1 = 1$.

The foregoing considerations clearly show the superiority of R_1 over R. In fact, as will be shown in part V, the interpretation of the results in terms of errors in coordinates becomes simple when one considers the normalized reliability index R_1. All these only point to the importance of treating the variables in their properly normalized forms.

References

ERDELYI, A. L. (1954). *Tables of Integral Transforms.* Vol. II, p. 307. New York: McGraw Hill.

RAMACHANDRAN, G. N. & SRINIVASAN, R. (1963). *Nature, Lond.* **200**, 1090.

RAMACHANDRAN, G. N., SRINIVASAN, R. & SARMA, V. R. (1963). *Acta Cryst.* **16**, 662.

SRINIVASAN, R., SARMA, V. R. & RAMACHANDRAN, G. N. (1963*a*). *Acta Cryst.* **16**, 1151.

SRINIVASAN, R., SARMA, V. R. & RAMACHANDRAN, G. N. (1963*b*). In *Crystallography & Crystal Perfection,* p. 85. Ed. G. N. Ramachandran. London: Academic Press.

SRINIVASAN, R., SUBRAMANIAN, E. & RAMACHANDRAN, G. N. (1964). *Acta Cryst.* **17**, 1010.

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

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