2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

notorious for its rather indeterminate mean and 'infinite' variance, resulting from the 'tail' of the denominator distributions extending through zero to negative values. The leading terms of the ratio distribution are given by Kendall & Stuart (1977, p. 288).

2.1.7. Non-ideal distributions: the correction-factor approach

2.1.7.1. *Introduction*

The probability density functions (p.d.f.'s) of the magnitude of the structure factor, presented in Section 2.1.5, are based on the central-limit theorem discussed above. In particular, the centric and acentric p.d.f.'s given by equations (2.1.5.11) and (2.1.5.8), respectively, are expected to account for the statistical properties of diffraction patterns obtained from crystals consisting of nearly equal atoms, which obey the fundamental assumptions of uniformity and independence of the atomic contributions and are not affected by noncrystallographic symmetry and dispersion. It is also assumed there that the number of atoms in the asymmetric unit is large. Distributions of structure-factor magnitudes which are based on the central-limit theorem, and thus obey the above assumptions, have been termed 'ideal', and the subjects of the following sections are those distributions for which some of the above assumptions/restrictions are not fulfilled; the latter distributions will be called 'non-ideal'.

We recall that the assumption of uniformity consists of the requirement that the fractional part of the scalar product $hx + ky + lz$ be uniformly distributed over the [0, 1] interval, which holds well if x, y, z are rationally independent (Hauptman & Karle, 1953), and permits one to regard the atomic contribution to the structure factor as a random variable. This is of course a necessary requirement for any statistical treatment. If, however, the atomic composition of the asymmetric unit is widely heterogeneous, the structure factor is then a sum of unequally distributed random variables and the Lindeberg–Lévy version of the central-limit theorem (*cf.* Section 2.1.4.4) cannot be expected to apply. Other versions of this theorem might still predict a normal p.d.f. of the sum, but at the expense of a correspondingly large number of terms/atoms. It is well known that atomic heterogeneity gives rise to severe deviations from ideal behaviour (*e.g.* Howells *et al.*, 1950) and one of the aims of crystallographic statistics has been the introduction of a correct dependence on the atomic composition into the non-ideal p.d.f.'s [for a review of the early work on non-ideal distributions see Srinivasan & Parthasarathy (1976)]. A somewhat less well known fact is that the dependence of the p.d.f.'s of $|E|$ on space-group symmetry becomes more conspicuous as the composition becomes more heterogeneous (*e.g.* Shmueli, 1979; Shmueli & Wilson, 1981). Hence both the composition and the symmetry dependence of the intensity statistics are of interest. Other problems, which likewise give rise to non-ideal p.d.f.'s, are the presence of heavy atoms in (variable) special positions, heterogeneous structures with complete or partial noncrystallographic symmetry, and the presence of outstandingly heavy dispersive scatterers.

The need for theoretical representations of non-ideal p.d.f.'s is exemplified in Fig. 2.1.7.1(*a*), which shows the ideal centric and acentric p.d.f.'s together with a frequency histogram of $|E|$ values, recalculated for a centrosymmetric structure containing a platinum atom in the asymmetric unit of $P\bar{1}$ (Faggiani *et al.*, 1980). Clearly, the deviation from the Gaussian p.d.f., predicted by the central-limit theorem, is here very large and a comparison with the possible ideal distributions can (in this case) lead to wrong conclusions.

Two general approaches have so far been employed in derivations of non-ideal p.d.f.'s which account for the above-mentioned problems: the correction-factor approach, to be dealt
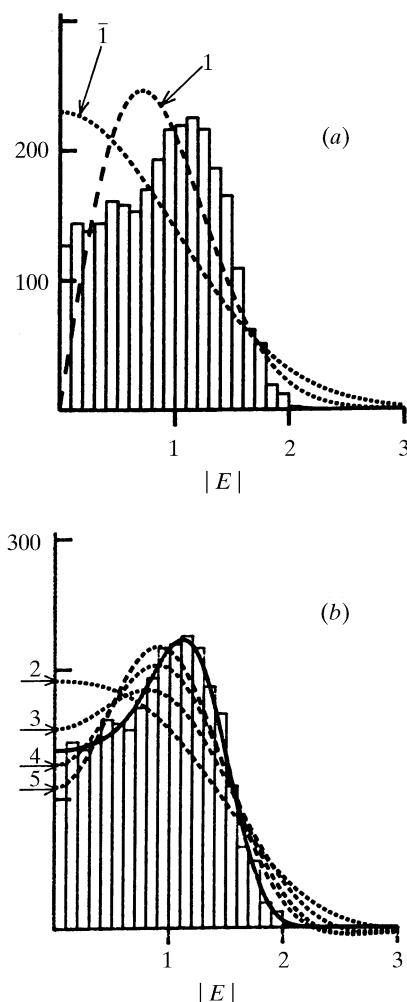

Fig. 2.1.7.1. Atomic heterogeneity and intensity statistics. The histogram appearing in (*a*) and (*b*) was constructed from $|E|$ values which were recalculated from atomic parameters published for the centrosymmetric structure of $C_6H_{18}Cl_2N_4O_4Pt$ (Faggiani *et al.*, 1980). The space group of the crystal is $P\bar{1}$, $Z = 2$, *i.e.* all the atoms are located in general positions. (*a*) A comparison of the recalculated distribution of $|E|$ with the ideal centric [equation (2.1.5.11)] and acentric [equation (2.1.5.8)] p.d.f.'s, denoted by $\bar{1}$ and 1, respectively. (*b*) The same recalculated histogram along with the centric correction-factor p.d.f. [equation (2.1.7.5)], truncated after two, three, four and five terms (dashed lines), and with that accurately computed for the correct space-group Fourier p.d.f. [equations (2.1.8.5) and (2.1.8.22)] (solid line).

with in the following sections, and the more recently introduced Fourier method, to which Section 2.1.8 is dedicated. In what follows, we introduce briefly the mathematical background of the correction-factor approach, apply this formalism to centric and acentric non-ideal p.d.f.'s, and present the numerical values of the moments of the trigonometric structure factor which permit an approximate evaluation of such p.d.f.'s for all the three-dimensional space groups.

2.1.7.2. *Mathematical background*

Suppose that $p(x)$ is a p.d.f. which accurately describes the experimental distribution of the random variable x, where x is related to a sum of random variables and can be assumed to obey (to some approximation) an ideal p.d.f., say $p^{(0)}(x)$, based on the central-limit theorem. In the correction-factor approach we seek to represent $p(x)$ as

$$p(x) = p^{(0)}(x) \sum_k d_k f_k(x), \qquad (2.1.7.1)$$

where d_k are coefficients which depend on the cause of the deviation of $p(x)$ from the central-limit theorem approximation and $f_k(x)$ are suitably chosen functions of x. A choice of the set $\{f_k\}$ is deemed suitable, if only from a practical point of view, if it allows the convenient introduction of the cause of the above deviation of $p(x)$ into the expansion coefficients d_k. This requirement is satisfied – also from a theoretical point of view – by taking $f_k(x)$ as a set of polynomials which are orthogonal with respect to the ideal p.d.f., taken as their weight function (*e.g.* Cramér, 1951). That is, the functions $f_k(x)$ so chosen have to obey the relationship

$$\int_a^b f_k(x) f_m(x) p^{(0)}(x)\, \mathrm{d}x = \delta_{km} = \begin{cases} 1, & \text{if } 2k = m \\ 0, & \text{if } 2k \neq m \end{cases}, \quad (2.1.7.2)$$

where $[a, b]$ is the range of existence of all the functions involved. It can be readily shown that the coefficients d_k are given by

$$d_k = \int_a^b f_k(x) p(x)\, \mathrm{d}x = \langle f_k(x) \rangle = \sum_{n=0}^k c_n^{(k)} \langle x^n \rangle, \qquad (2.1.7.3)$$

where the brackets $\langle\ \rangle$ in equation (2.1.7.3) denote averaging with respect to the *unknown* p.d.f. $p(x)$ and $c_n^{(k)}$ is the coefficient of the nth power of x in the polynomial $f_k(x)$. The coefficients d_k are thus directly related to the moments of the non-ideal distribution and the coefficients of the powers of x in the orthogonal polynomials. The latter coefficients can be obtained by the Gram–Schmidt procedure (*e.g.* Spiegel, 1974), or by direct use of the Szegö determinants (*e.g.* Cramér, 1951), for any weight function that has finite moments. However, the feasibility of the present approach depends on our ability to obtain the moments $\langle x^n \rangle$ *without* the knowledge of the non-ideal p.d.f., $p(x)$.

2.1.7.3. *Application to centric and acentric distributions*

We shall summarize here the non-ideal centric and acentric distributions of the magnitude of the normalized structure factor E (*e.g.* Shmueli & Wilson, 1981; Shmueli, 1982). We assume that (i) all the atoms are located in general positions and have rationally independent coordinates, (ii) all the scatterers are dispersionless, and (iii) there is no noncrystallographic symmetry. Arbitrary atomic composition and space-group symmetry are admitted. The appropriate weight functions and the corresonding orthogonal polynomials are

$p^{(0)}(\lvert E \rvert)$	$f_k(x)$	Non-ideal distribution
$(2/\pi)^{1/2} \exp(-\lvert E \rvert^2/2)$	$He_{2k}(\lvert E \rvert)/[(2k)!]^{1/2}$	Centric
$2\lvert E \rvert \exp(-\lvert E \rvert^2)$	$L_k(\lvert E \rvert^2)$	Acentric

$$(2.1.7.4)$$

where He_k and L_k are Hermite and Laguerre polynomials, respectively, as defined, for example, by Abramowitz & Stegun (1972). Equations (2.1.7.2), (2.1.7.3) and (2.1.7.4) suffice for the general formulation of the above non-ideal p.d.f.'s of $\lvert E \rvert$. Their full derivation entails (i) the expression of a sufficient number of moments of $\lvert E \rvert$ in terms of absolute moments of the trigonometric structure factor (*e.g.* Shmueli & Wilson, 1981; Shmueli, 1982) and (ii) calculation of the latter moments for the various symmetries (Wilson, 1978*b*; Shmueli & Kaldor, 1981, 1983). The notation below is similar to that employed by Shmueli (1982).

These non-ideal p.d.f.'s of $\lvert E \rvert$, for which the first five expansion terms are available, are given by

$$p_c(\lvert E \rvert) = p_c^{(0)}(\lvert E \rvert) \left[1 + \sum_{k=2}^{\infty} \frac{A_{2k}}{(2k)!} He_{2k}(\lvert E \rvert) \right] \qquad (2.1.7.5)$$

and

$$p_a(\lvert E \rvert) = p_a^{(0)}(\lvert E \rvert) \left[1 + \sum_{k=2}^{\infty} \frac{(-1)^k B_{2k}}{k!} L_k(\lvert E \rvert^2) \right] \qquad (2.1.7.6)$$

for centrosymmetric and noncentrosymmetric space groups, respectively, where $p_c^{(0)}(\lvert E \rvert)$ and $p_a^{(0)}(\lvert E \rvert)$ are the ideal centric and acentric p.d.f.'s [see (2.1.7.4)] and the unified form of the coefficients A_{2k} and B_{2k}, for $k = 2, 3, 4$ and 5, is

$$
\begin{aligned}
A_4 \quad \text{or} \quad & B_4 = a_4 Q_4 \\
A_6 \quad \text{or} \quad & B_6 = a_6 Q_6 \\
A_8 \quad \text{or} \quad & B_8 = a_8 Q_8 + U(a_4^2 Q_4^2 - \gamma_4^2) \\
A_{10} \quad \text{or} \quad & B_{10} = a_{10} Q_{10} + V(a_4 a_6 Q_4 Q_6 - \gamma_4 \gamma_6 Q_{10}) \\
& \qquad + W \gamma_4^2 Q_{10}
\end{aligned}
\qquad (2.1.7.7)
$$

(Shmueli, 1982), where $U = 35$ or 18, $V = 210$ or 100 and $W = 3150$ or 900 according as A_{2k} or B_{2k} is required, respectively, and the other quantities in equation (2.1.7.7) are given below. The *composition-dependent* terms in equations (2.1.7.7) are

$$Q_{2k} = \frac{\sum_{j=1}^m f_j^{2k}}{\left(\sum_{n=1}^m f_n^2\right)^k}, \qquad (2.1.7.8)$$

where m is the number of atoms in the asymmetric unit, f_j, $j = 1, \ldots, m$ are their scattering factors, and the *symmetry* dependence is expressed by the coefficients a_{2k} in equation (2.1.7.7), as follows:

$$a_{2k} = (-1)^{k-1}(k-1)! \alpha_{k0} + \sum_{p=2}^k (-1)^{k-p}(k-p)! \alpha_{kp} \gamma_{2p},$$

$$(2.1.7.9)$$

where

$$\alpha_{kp} = \binom{k}{p} \frac{(2k-1)!!}{(2p-1)!!} \quad \text{or} \quad \binom{k}{p} \frac{k!}{p!} \qquad (2.1.7.10)$$

according as the space group is centrosymmetric or noncentrosymmetric, respectively, and γ_{2p} in equation (2.1.7.9) is given by

$$\gamma_{2p} = \frac{\langle \lvert T \rvert^{2p} \rangle}{\langle \lvert T \rvert^2 \rangle^p}, \qquad (2.1.7.11)$$

where $\langle \lvert T \rvert^k \rangle$ is the kth absolute moment of the trigonometric structure factor

$$T(\mathbf{h}) = \sum_{s=1}^g \exp[2\pi i \mathbf{h}^T (\mathbf{P}_s \mathbf{r} + \mathbf{t}_s)] \equiv \xi(\mathbf{h}) + i\eta(\mathbf{h}). \qquad (2.1.7.12)$$

In equation (2.1.7.12), g is the number of general equivalent positions listed in *IT* A (1983) for the space group in question, times the multiplicity of the Bravais lattice, $(\mathbf{P}_s, \mathbf{t}_s)$ is the sth space-group operator and $\mathbf{r}$ is an atomic position vector.