International Tables for Crystallography (2006). Vol. B, Section 2.1.4.5, p. 195.

Surprisingly, this theorem has a very wide applicability and values of n as low as 30 are often large enough for the theorem to be useful. Situations in which the normal p.d.f. must be modified or replaced by an altogether different one are dealt with in Sections 2.1.7 and 2.1.8 of this chapter.

2.1.4.4. *Conditions of validity*

The above outline of a proof of the central-limit theorem depended on the existence of moments of all orders. The components of structure factors always possess finite moments of all orders, but the existence of moments beyond the second is not necessary for the validity of the theorem and it can be proved under much less stringent conditions. In fact, if all the random variables in equation (2.1.4.19) have the *same* distribution – as in a homoatomic structure – the only requirement is that the second moments of the distributions should exist [the Lindeberg–Lévy theorem (*e.g.* Cramér, 1951)]. If the distributions are not the same – as in a heteroatomic structure – some further condition is necessary to ensure that no individual random variable dominates the sum. The Liapounoff proof requires the existence of third absolute moments, but this is regarded as aesthetically displeasing; a theorem that ultimately involves only means and variances should require only means and variances in the proof. The Lindeberg–Cramér conditions meet this aesthetic criterion. Roughly, the conditions are that S^2, the variance of the sum, should tend to infinity and σ_j^2/S^2, where σ_j^2 is the variance of the jth random variable, should tend to zero for all j as n tends to infinity. The precise formulation is quoted by Kendall & Stuart (1977, p. 207).

2.1.4.5. *Non-independent variables*

The central-limit theorem, under certain conditions, remains valid even when the variables summed in equation (2.1.4.19) are not independent. The conditions have been investigated by Bernstein (1922, 1927); roughly they amount to requiring that the variables should not be too closely correlated. The theorem applies, in particular, when each x_r is related to a finite number, $f(n)$, of its neighbours, when the x's are said to be $f(n)$ dependent. The $f(n)$ dependence seems plausible for crystallographic applications, since the positions of atoms close together in a structure are closely correlated by interatomic forces, whereas those far apart will show little correlation if there is any flexibility in the asymmetric unit when unconstrained. Harker's (1953) idea of 'globs' seems equivalent to $f(n)$ dependence. Long-range stereochemical effects, as in pseudo-graphitic aromatic hydrocarbons, would presumably produce long-range correlations and make $f(n)$ dependence less plausible. If Bernstein's conditions are satisfied, the central-limit theorem would apply, but the actual value of $\langle x^2\rangle - \langle x\rangle^2$ would have to be used for the variance, instead of the sum of the variances of the random variables in (2.1.4.19). Because of the correlations the two values are no longer equal.

French & Wilson (1978) seem to have been the first to appeal explicitly to the central-limit theorem extended to non-independent variables, but many previous workers [for typical references, see Wilson (1981)] tacitly made the replacement – in the X-ray case substituting the local mean intensity for the sum of the squares of the atomic scattering factors.

2.1.5. Ideal probability density distributions

In applications of the central-limit theorem, and its extensions, to intensity statistics the x_j's of equation (2.1.4.19) have the form (atomic scattering factor of the jth atom) times (a trigonometric expression characteristic of the space group and Wyckoff position; also known as the trigonometric structure factor). These trigonometric expressions for all the space groups, and general Wyckoff positions, are given in Tables A1.4.3.1 through A1.4.3.7, and their first few even moments (fixed-index averaging) are given in Table 2.1.7.1. One cannot, of course, conclude that the magnitudes of the structure factor always have a normal distribution – even if the structure is homoatomic; one must look at each problem and see what components of the structure factor can be put in the form (2.1.4.19), deduce the m and σ^2 to be used for each, and combine the components to obtain the asymptotic (large N, not large x) expression for the problem in question. Ordinarily the components are the real and the imaginary parts of the structure factor; the structure factor is purely real only if the structure is centrosymmetric, the space-group origin is chosen at a crystallographic centre and the atoms are non-dispersive.

2.1.5.1. *Ideal acentric distributions*

The ideal acentric distributions are obtained by applying the central-limit theorem to the real and the imaginary parts of the structure factor, as given by equation (2.1.1.1). Consider first a crystal with no rotational symmetry (space group $P1$). The real part, A, of the structure factor is then given by

$$A = \sum_{j=1}^{N} f_j \cos\vartheta_j, \qquad (2.1.5.1)$$

where N is the number of atoms in the unit cell and ϑ_j is the phase angle of the jth atom. The central-limit theorem then states that A tends to be normally distributed about its mean value with variance equal to its mean-square deviation from its mean. Under the assumption that the phase angles ϑ_j are uniformly distributed on the 0–2π range, the mean value of each cosine is zero, so that its variance is

$$\sigma^2 = \sum_{j=1}^{N} f_j^2 \langle\cos^2\vartheta_j\rangle. \qquad (2.1.5.2)$$

Under the same assumption, the mean value of each $\cos^2\vartheta$ is one-half, so that the variance becomes

$$\sigma^2 = (1/2)\sum_{j=1}^{N} f_j^2 = (1/2)\Sigma, \qquad (2.1.5.3)$$

where Σ is the sum of the squares of the atomic scattering factors [*cf.* equation (2.1.2.4)]. The asymptotic form of the distribution of A is therefore given by

$$p(A)\,\mathrm{d}A = (\pi\Sigma)^{-1/2}\exp(-A^2/\Sigma)\,\mathrm{d}A. \qquad (2.1.5.4)$$

A similar calculation, with sines instead of cosines, gives an analogous distribution for the imaginary part B, so that the joint probability of the real and imaginary parts of F is

$$p(A,B)\,\mathrm{d}A\,\mathrm{d}B = (\pi\Sigma)^{-1}\exp[-(A^2+B^2)/\Sigma]\,\mathrm{d}A\,\mathrm{d}B. \qquad (2.1.5.5)$$

Ordinarily, however, we are more interested in the distribution of the magnitude, $|F|$, of the structure factor than in the distribution of A and B. Using polar coordinates in equation (2.1.5.5) [$A = |F|\cos\phi$, $B = |F|\sin\phi$] and integrating over the angle ϕ gives

$$p(|F|)\,\mathrm{d}|F| = (2|F|/\Sigma)\exp(-|F|^2/\Sigma)\,\mathrm{d}|F|. \qquad (2.1.5.6)$$

It is usually convenient, in structure-factor and intensity statistics, to express the results in terms of the normalized structure factor E and its magnitude $|E|$. If $|F|$ has been put on an absolute scale (see Section 2.2.4.3), we have

$$E = \frac{F}{\sqrt{\Sigma}} \quad\text{and}\quad |E| = \frac{|F|}{\sqrt{\Sigma}}, \qquad (2.1.5.7)$$

Copyright © 2006 International Union of Crystallography

references