SAYARI, A. & BILLIET, Y. (1977). *Acta Cryst.* A33, 985-986.
SAYARI, A., BILLIET, Y. & ZARROUK, H. (1978). *Acta Cryst.* A34, 553-555.
SEITZ, F. (1935*a*). *Z. Kristallogr.* 90, 289-313.
SEITZ, F. (1935*b*). *Z. Kristallogr.* 91, 336-366.
WOLFF, P. M. DE, BILLIET, Y., DONNAY, J. D. H., FISCHER, W., GALIULIN, R. B., GLAZER, A. M., SENECHAL, M., SHOEMAKER, D. P., WONDRATSCHEK, H., HAHN, TH., WILSON, A. J. C. & ABRAHAMS, S. C. (1989). *Acta Cryst.* A45, 494-499.

Acta Cryst. (1990). A46, 552-559

Intensity Distributions in Fiber Diffraction

BY R. P. MILLANE

The Whistler Center for Carbohydrate Research, Smith Hall, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 5 *August* 1989; *accepted* 20 *December* 1989)

Abstract

The probability distributions of X-ray intensities in fiber diffraction are different from those for single crystals (Wilson statistics) because of the cylindrical averaging of the diffraction data. Stubbs [*Acta Cryst.* (1989), A45, 254-258] has recently determined the intensity distributions on a fiber diffraction pattern for a fixed number of overlapping Fourier-Bessel terms. Some properties of the amplitude and intensity distributions are derived here. It is shown that the amplitudes and intensities are approximately normally distributed (the distributions being asymptotically normal with increasing number of Fourier-Bessel terms). Improved approximations using an Edgeworth series are derived. Other statistical properties and some asymptotic expansions are also derived, and normalization of fiber diffraction amplitudes is discussed. The accuracies of the normal approximations are illustrated for particular fiber structures, and possible applications of intensity statistics in fiber diffraction are discussed.

Notation

$\mathcal{G}(\mathcal{I})$	amplitude (intensity) on a fiber diffraction pattern.
$\mathcal{E}$	normalized amplitude.
m	number of degrees of freedom for $\mathcal{G}$.
$P_m(\mathcal{G}), P_m(\mathcal{I})$	probability density functions for $\mathcal{G}$ and $\mathcal{I}$.
α_{mn} (β_{mn})	nth moment of $\mathcal{G}(\mathcal{I})$.
μ_m (ν_m)	mean of $\mathcal{G}(\mathcal{I})$.
σ_m^2 (τ_m^2)	variance of $\mathcal{G}(\mathcal{I})$.
μ_{mn} (ν_{mn})	nth central moment of $\mathcal{G}(\mathcal{I})$.
$Q_m(\mathcal{G}), Q_m(\mathcal{I})$	cumulative distribution functions for $\mathcal{G}$ and $\mathcal{I}$.
$\varphi_m(y), \psi_m(y)$	characteristic functions for $\mathcal{G}$ and $\mathcal{I}$.
κ_{mn}	nth cumulant for $\mathcal{I}$.
$\hat{P}_m(\mathcal{G}), \hat{P}'_m(\mathcal{G}), \hat{P}_m(\mathcal{I})$	normal approximations to $P_m(\mathcal{G})$ and $P_m(\mathcal{I})$.
$\tilde{P}_m(\mathcal{G}), \tilde{P}'_m(\mathcal{G}), \tilde{P}_m(\mathcal{I})$	Edgeworth series approximations to $P_m(\mathcal{G})$ and $P_m(\mathcal{I})$.
$\hat{Q}_m(\mathcal{G}), \hat{Q}'_m(\mathcal{G}), \hat{Q}_m(\mathcal{I})$	normal approximations to $Q_m(\mathcal{G})$ and $Q_m(\mathcal{I})$.
$\hat{\varphi}_m(y), \hat{\psi}_m(y)$	normal approximations to $\varphi_m(y)$ and $\psi_m(y)$.

1. Introduction

Statistical descriptions of X-ray amplitudes have played important roles in many aspects of crystallography. The most remarkable, of course, is the use of conditional distributions of phases in direct methods for phase determination (Hauptman & Karle, 1953; Giacovazzo, 1980; Bricogne, 1984). Other applications include detection of symmetry (Wilson, 1949), analysis of twinning (Yeates, 1988), and estimation of R factors (Wilson, 1950; Luzatti, 1952). The initial application of such ideas was a study of the distribution of intensities diffracted by a crystal (Wilson, 1949).

X-ray fiber diffraction is a variant of traditional crystallography that can be used to determine structures of molecules that prefer to form fibers rather than single crystals (Millane, 1988). In a fiber specimen, the diffracting particles are randomly rotated so that the diffraction pattern is cylindrically averaged. Intensity distributions in fiber diffraction are therefore different from those in traditional crystallography. Although intensity statistics have not yet been utilized in fiber diffraction, it may be possible to develop useful applications. The first step in this

0108-7673/90/070552-08$03.00

direction was taken by Stubbs (1989) who derived the distribution of amplitudes on a fiber diffraction pattern as a function of the number of overlapping Fourier-Bessel structure factors. This is essentially an extension of Wilson statistics to the fiber diffraction case. These results have been used to estimate largest likely R factors in fiber diffraction analyses (Stubbs, 1989; Millane, 1989*a*, *b*, 1990).

Some properties of the distributions of amplitudes (and intensities) on a fiber diffraction pattern are examined here. In particular, it is shown that the diffracted amplitudes and intensities are approximately normally distributed. Some other statistical parameters and asymptotic properties are also derived. The insights and simplifications afforded by normal distributions, as well as the other properties derived, may be useful in developing applications of intensity statistics in fiber diffraction.

Wilson statistics, relevant aspects of fiber diffraction theory and intensity distributions in fiber diffraction derived by Stubbs (1989) are briefly reviewed in the next section. Properties of the distributions of intensities and amplitudes in fiber diffraction are derived in §§ 3 and 4 respectively. Possible normalizations of fiber diffraction amplitudes are discussed in the next section. The accuracy of approximate normal distributions for typical fiber diffraction problems is discussed in § 6. Possible applications of this work and concluding remarks are made in the final section.

2. Preliminaries

Wilson (1949) showed that the probability density function for structure amplitudes $P(F)$ diffracted by a centric crystal is

$$P(F) = (2/\pi\varepsilon)^{1/2} \exp(-F^2/2\varepsilon) \tag{1}$$

and the corresponding density for the intensities, $P(I)$, is therefore

$$P(I) = (2\pi\varepsilon)^{-1/2} I^{-1/2} \exp(-I/2\varepsilon) \tag{2}$$

where

$$\varepsilon = \sum_j f_j^2 \tag{3}$$

and the f_j are the atomic scattering factors. Note that the coefficient in (1) is incorrect in the *Abstract* of Wilson (1949). The densities for a non-centric crystal are

$$P(F) = (2/\varepsilon) F \exp(-F^2/\varepsilon) \tag{4}$$

$$P(I) = \varepsilon^{-1} \exp(-I/\varepsilon). \tag{5}$$

These distributions apply in a thin resolution shell where ε is constant.

In fiber diffraction, the diffracting particles are randomly rotated about their long axes and the diffracted intensity $I_l(R)$ at a reciprocal-space cylindrical radius R on layer line l is given by (Franklin & Klug, 1955)

$$I_l(R) = \sum_n |G_{nl}(R)|^2 \tag{6}$$

where the sum includes only those n that satisfy the helix selection rule (Cochran, Crick & Vand, 1952), the $G_{nl}(R)$ (sometimes abbreviated to G_n) are the complex Fourier-Bessel structure factors (Klug, Crick & Wyckoff, 1958) given by

$$G_{nl}(R) = \sum_j f_j J_n(2\pi R r_j) \exp[i(-n\varphi_j + 2\pi l z_j/c)], \tag{7}$$

$J_n(x)$ is the nth-order Bessel function of the first kind, and (r_j, φ_j, z_j) are the cylindrical polar coordinates of the jth atom. Although the sum in (6) is in principle infinite, it is in practice finite as a result of the behavior of Bessel functions (Makowski, 1982). It is convenient to define an m-dimensional vector $\mathcal{G}$ whose components are the real and imaginary parts of the G_n terms (of significant value) that contribute to the diffracted intensity at a particular position in reciprocal space (Namba & Stubbs, 1987). In general, therefore, m is twice the number of terms, but if any terms are real, m will be less than this (Stubbs, 1989; Millane 1989*a*). The quantity m is sometimes referred to as the number of degrees of freedom of $\mathcal{G}$. The measured amplitude at a particular position in reciprocal space is therefore equal to the length $\mathcal{G}$ of $\mathcal{G}$ and the intensity $\mathcal{I}$ is $\mathcal{I} = \mathcal{G}^2$, or

$$\mathcal{I} = \sum_{i=1}^{m} \mathcal{G}_i^2 = \sum_{i=1}^{m/2} \mathcal{I}_i \tag{8}$$

where the $\mathcal{G}_i$ are the components of $\mathcal{G}$ and the $\mathcal{I}_i$ are the intensities that contribute to $\mathcal{I}$. Stubbs (1989) showed that, for a particular value of m, the probability density functions for $\mathcal{G}$ and $\mathcal{I}$ are given by [referring to Millane (1989*a*) and utilizing the gamma function $\Gamma(x)$]

$$P_m(\mathcal{G}) = [2\varepsilon^{-m/2}/\Gamma(m/2)] \mathcal{G}^{m-1} \exp(-\mathcal{G}^2/\varepsilon) \tag{9}$$

$$P_m(\mathcal{I}) = [\varepsilon^{-m/2}/\Gamma(m/2)] \mathcal{I}^{m/2-1} \exp(-\mathcal{I}/\varepsilon) \tag{10}$$

where

$$\varepsilon = \sum_j f_j^2 J_n^2(2\pi R r_j). \tag{11}$$

$P_m(\mathcal{G})$ and $P_m(\mathcal{I})$ are shown as the solid lines in Figs. 1 and 2 for some typical values of m (for all examples in this paper $\varepsilon = 1$).

Comparison of (9) and (10) with (1)-(5) shows that the $m = 2$ case is identical (except for the definition of ε) to the non-centrosymmetric single-crystal case, and the $m = 1$ case is similar, although not identical, to the centrosymmetric case. The latter difference is due to a halving of the number of independent atoms for a centrosymmetric crystal that does not occur in the fiber diffraction case.

3. Intensity distributions

The density $P_m(\mathcal{I})$ is a gamma or χ^2 type of distribution (Abramowitz & Stegun, 1972, chap. 26). Straightforward calculations show that the nth moment β_{mn} is given by

$$\beta_{mn} = \Gamma(m/2+n)/\Gamma(m/2). \tag{12}$$

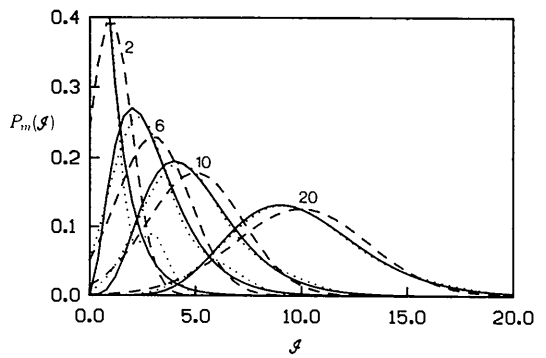

Fig. 1. Probability density functions for the intensity on a fiber diffraction pattern for different values of m. The numbers adjacent to the curves indicate the value of m. The different curves are the exact densities $P_m(\mathcal{I})$ (——), normal approximations $\hat{P}_m(\mathcal{I})$ (- - -) and Edgeworth series approximations $\tilde{P}_m(\mathcal{I})$ (· · ·).

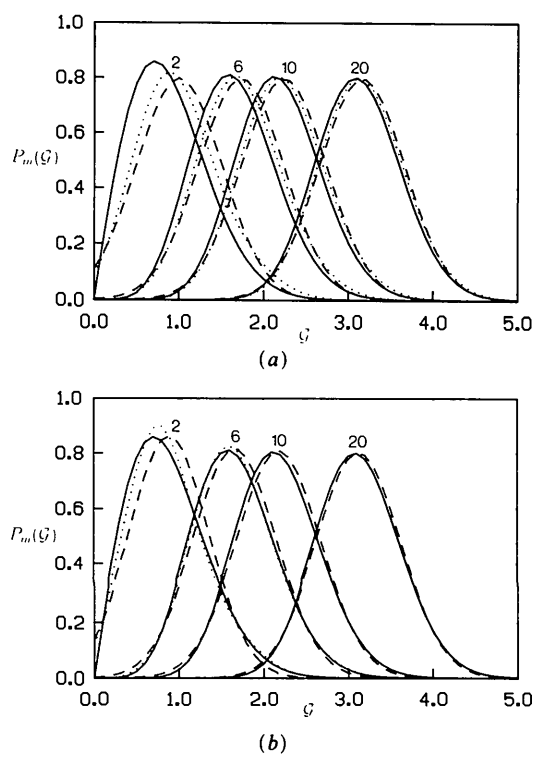

Fig. 2. Probability density functions for the amplitude on a fiber diffraction pattern for different values of m. (a) Exact densities $P_m(\mathcal{G})$ (——), normal approximations $\hat{P}_m(\mathcal{G})$ (- - -) and Edgeworth series approximations $\tilde{P}_m(\mathcal{G})$ (· · ·). (b) Curves are the same as in (a) except that the more accurate approximations $\hat{P}'_m(\mathcal{G})$ and $\tilde{P}'_m(\mathcal{G})$ are used.

The mean $\nu_m = \beta_{m1}$, variance $\tau_m^2 = \nu_{m2}$, and third central moment, ν_{m3}, are given by

$$\nu_m = \varepsilon m/2 \tag{13}$$

$$\tau_m^2 = \varepsilon^2 m/2 \tag{14}$$

and

$$\nu_{m3} = \varepsilon^3 m. \tag{15}$$

The cumulative distribution function $Q_m(\mathcal{I})$ is given by

$$Q_m(\mathcal{I}) = [1/\Gamma(m/2)]\gamma(m/2, \mathcal{I}/\varepsilon) \tag{16}$$

where $\gamma(a, x)$ is the incomplete gamma function (Abramowitz & Stegun, 1972, equation 6.5.2), and is shown as the solid lines in Fig. 3. With the standard results for gamma distributions (Abramowitz & Stegun, 1972, chap. 26), the characteristic function $\psi_m(y)$ and nth cumulant κ_{mn} for $\mathcal{I}$ are given by

$$\psi_m(y) = (1 - i\varepsilon y)^{-m/2} \tag{17}$$

and

$$\kappa_{mn} = \varepsilon^n m\Gamma(n)/2. \tag{18}$$

Inspection of (8), (13) and (14) shows that $\mathcal{I}$ is equal to the sum of $m/2$ identically distributed random variables ($\mathcal{I}_i$) with mean ε and variance ε^2. Therefore, by the central limit theorem, the probability density $P_m(\mathcal{I})$ is asymptotically normal as $m \to \infty$. For finite m, $P_m(\mathcal{I})$ is then approximately normal with mean $\varepsilon m/2$ and variance $\varepsilon^2 m/2$. Denoting this approximation by $\hat{P}_m(\mathcal{I})$ we obtain

$$\hat{P}_m(\mathcal{I}) = (\pi m)^{-1/2}\varepsilon^{-1} \exp[-(\mathcal{I} - \varepsilon m/2)^2/(\varepsilon^2 m)]. \tag{19}$$

The approximation is compared with the exact density function in Fig. 1 and the maximum error is about 0·05 for $m > 10$. Asymptotic corrections to the normal approximation can be obtained by developing an Edgeworth series (Klug, 1958; Cramer, 1970) for $P(\mathcal{I})$. Since subsequent terms in such a series become

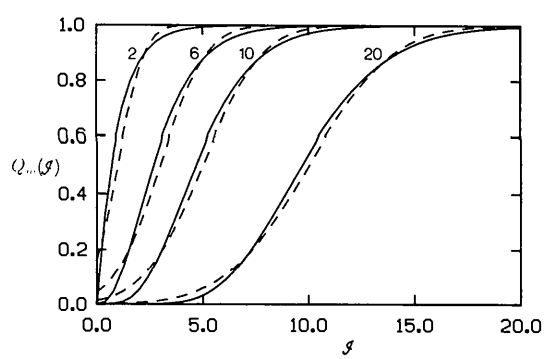

Fig. 3. Cumulative distribution functions for intensities on a fiber diffraction pattern for different values of m. The different curves are the exact distributions $Q_m(\mathcal{I})$ (——) and the normal approximations $\hat{Q}_m(\mathcal{I})$ (- - -).

increasingly complicated, only the first term is given here. Use of ν_{m3} given by (15) with the corrected ('Edgeworth series') density denoted by $\tilde{P}_m(\mathcal{I})$ gives

$$\tilde{P}_m(\mathcal{I}) = \hat{P}_m(\mathcal{I})\{1+(\sqrt{2}/3)m^{-1/2} \times H_3[\varepsilon^{-1}(m/2)^{-1/2}(\mathcal{I}-\varepsilon m/2)]\} \quad (20)$$

where $H_3(x)=x^3-3x$ is the third-order Hermite polynomial. This approximation is shown as the dotted line in Fig. 1 and is seen to be a significant improvement over the normal approximation, the maximum error being about 0·05 for $m>6$.

The normal approximation (19) can be used to derive an approximate normal cumulative distribution function $\hat{Q}_m(\mathcal{I})$ for $\mathcal{I}$ given by

$$\hat{Q}_m(\mathcal{I}) = (1/2)\{1+\text{erf}[(\mathcal{I}-\varepsilon m/2)/(\varepsilon m^{1/2})]\} \quad (21)$$

where erf(x) is the error function. $\hat{Q}_m(\mathcal{I})$ is compared with $Q_m(\mathcal{I})$ in Fig. 3 and is seen to be a good approximation. An improved approximation could be derived from the Edgeworth series (20) but its complexity compared with (16) would make it of little use. The normal approximation (19) can also be used to derive an approximation $\hat{\psi}_m(y)$ to the characteristic function given by

$$\hat{\psi}_m(y) = \exp(i\varepsilon m y/2 - \varepsilon^2 m y^2/4). \quad (22)$$

4. Amplitude distributions

Using (9) and standard integrals (Gradshteyn & Ryzhik, 1980, equation 3.461), we may easily show that the nth moment α_{mn} of $\mathcal{G}$ is given by

$$\alpha_{mn} = \varepsilon^{n/2}\Gamma(m/2+n/2)/\Gamma(m/2) \quad (23)$$

so that the mean $\mu_m = \alpha_{m1}$ is

$$\mu_m = \varepsilon^{1/2}\Gamma(m/2+1/2)/\Gamma(m/2). \quad (24)$$

Using these results, we can write the variance and third central moment in the form

$$\sigma_m^2 = (\varepsilon m/2) - \mu_m^2 \quad (25)$$

and

$$\mu_{m3} = \varepsilon[(1/2)-m]\mu_m + 2\mu_m^3. \quad (26)$$

Use of Stirling's expansion for the gamma function (Abramowitz & Stegun, 1972, equation 6.1.37) shows that the asymptotic behavior of the mean is

$$\mu_m = (\varepsilon m/2)^{1/2}[1-(1/4)m^{-1} -(1/8)m^{-2}+O(m^{-3})] \quad m\to\infty \quad (27)$$

where $O(x)$ denotes terms of order x and sufficient terms are given to obtain the leading term for the third central moment given below. Substitution of (27) into (25) and (26) gives the asymptotic expansions for σ_m^2 and μ_{m3} as

$$\sigma_m^2 = (\varepsilon/4)[1+(3/8)m^{-1}+O(m^{-2})] \quad m\to\infty \quad (28)$$

and

$$\mu_{m3} = (1/8)2^{-1/2}\varepsilon^{3/2}m^{-1/2}[1+O(m^{-1})] \quad m\to\infty. \quad (29)$$

The exact values for μ_m, σ_m^2 and μ_{m3}, together with the leading term of their asymptotic expansions, are shown in Fig. 4. The leading terms are seen to be quite accurate, except in the case of the variance for very small values of m.

Straightforward calculation shows that the cumulative distribution function for $\mathcal{G}$ is

$$Q_m(\mathcal{G}) = [1/\Gamma(m/2)]\gamma(m/2, \mathcal{G}^2/\varepsilon) \quad (30)$$

which is shown as the solid lines in Fig. 5. The characteristic function for $\mathcal{G}$, $\varphi_m(y)$ can be calculated using a standard integral (Gradshteyn & Ryzhik, 1980, equation 3.462.1) followed by application of the duplication formula for the gamma function (Abramowitz & Stegun, 1972, equation 6.1.18) giving

$$\varphi_m(y) = 2^{m/2}\pi^{-1/2}\Gamma(m/2+1/2) \times D_{-m}[-i(\varepsilon/2)^{1/2}y]\exp(-\varepsilon y^2/8) \quad (31)$$

where $D_m(x)$ is the parabolic cylinder function (Abramowitz & Stegun, 1972, chap. 19).

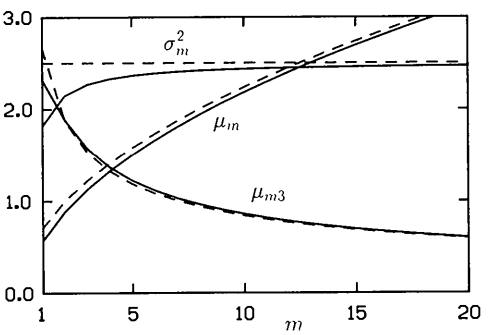

Fig. 4. Mean, variance (×10) and third central moment (×30) of $\mathcal{G}$ (——) as a function of m together with their leading-order asymptotic expansions (- - -).

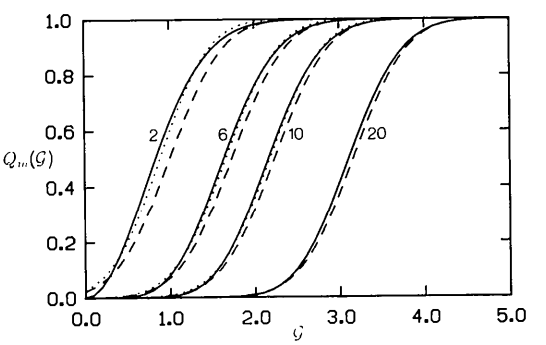

Fig. 5. Cumulative distribution functions for amplitudes on a fiber diffraction pattern for different values of m. The different curves are the exact distributions $Q_m(\mathcal{G})$ (——), and the normal approximations $\hat{Q}_m(\mathcal{G})$ (- - -) and $\hat{Q}'_m(\mathcal{G})$ (· · ·).

Since $\mathscr{G}$ is not a sum of random variables, the central limit cannot be applied to its distribution directly. However, it can be shown (Appendix A) that $P(\mathscr{G})$ is asymptotically normal as $m\to\infty$ with mean $(\varepsilon m/2)^{1/2}$ and variance $\varepsilon/4$. The normal approximation, denoted by $\hat{P}_m(\mathscr{G})$, is therefore given by

$$\hat{P}_m(\mathscr{G})=2(2/\pi)^{1/2}\varepsilon^{-1}\exp\{-2[\mathscr{G}-(\varepsilon m/2)^{1/2}]^2/\varepsilon\} \qquad m\to\infty. \quad (32)$$

This approximation is compared with the exact distribution in Fig. 2(a), and the maximum error is about 0·14 for $m>6$. Analysis of the asymptotic behavior of the third moment (Appendix A) shows that a correction to (32) in the form of the next term in the Edgeworth series is given by

$$\tilde{P}_m(\mathscr{G})=\hat{P}_m(\mathscr{G})(1+(\sqrt{2}/3)m^{-1/2}\times H_3\{2\varepsilon^{-1/2}[\mathscr{G}-(\varepsilon m/2)^{1/2}]\}). \quad (33)$$

This approximation is shown as the dotted lines in Fig. 2(a) and is seen to give a significant improvement over $\hat{P}_m(\mathscr{G})$, the maximum error being about 0·1 for $m>6$.

The mean, variance and third central moment of the approximate densities given in (32) and (33) are the leading terms in their asymptotic expansions (27)–(29). At the cost of a slight increase in complexity, therefore, a more accurate normal density function for the amplitude is

$$\hat{P}'_m(\mathscr{G})=(2\pi)^{-1/2}\sigma_m^{-1}\exp[-(\mathscr{G}-\mu_m)^2/2\sigma_m^2] \quad (34)$$

where μ_m and σ_m are given by (24) and (25) respectively. Similarly, from the exact expression (26) for the third central moment, a correction to (34) is given by

$$\tilde{P}'_m(\mathscr{G})=\hat{P}'_m(\mathscr{G})\{1+(\mu_{m3}/6\sigma_m^3)\times H_3[(\mathscr{G}-\mu_m)/\sigma_m]\}. \quad (35)$$

These two approximations are shown in Fig. 2(b) and are improvements on the previous approximations (Fig. 2a), the maximum errors being about 0·07 and 0·02 for $\hat{P}'_m(\mathscr{G})$ and $\tilde{P}'_m(\mathscr{G})$, respectively.

Using the Gaussian density function (32), we can obtain an approximate normal cumulative distribution function for $\mathscr{G}$ as

$$\hat{Q}_m(\mathscr{G})=(1/2)(1+\mathrm{erf}\{\sqrt{2}\varepsilon^{-1/2}[\mathscr{G}-(\varepsilon m/2)^{1/2}]\}), \quad (36)$$

and a slightly more complicated expression, denoted by $\hat{Q}'_m(\mathscr{G})$, can be obtained by using (34) instead of (32). These approximations are shown in Fig. 5 and are seen to be quite accurate. The normal approximation (32) can also be used to derive an approximate characteristic function $\hat{\varphi}_m(y)$ given by

$$\hat{\varphi}_m(y)=\exp[i(\varepsilon m/2)^{1/2}y-\varepsilon y^2/8]. \quad (37)$$

It can be shown (Appendix B) that $\hat{\varphi}_m(y)$ is the first term in the asymptotic expansion of $\varphi_m(y)$ as $m\to\infty$. A more accurate approximate characteristic function could be derived by using (34) instead of (32).

5. Normalization

It is common in crystallography to use normalized structure factors $E_{hkl}=F_{hkl}/\langle F_{hkl}^2\rangle^{1/2}$ and intensities $|E_{hkl}|^2=I_{hkl}/\langle I_{hkl}\rangle$ (Giacovazzo, 1980, chap. 1). This takes into account systematic variations in the structure factors with resolution (through ε), and in particular zones. Similar normalizations may also be useful in fiber diffraction and are discussed here.

The effect of ε can be removed by defining a normalized amplitude $\mathscr{E}$ by $\mathscr{E}=\mathscr{G}/\varepsilon^{1/2}$ and a normalized intensity by $\mathscr{E}^2=\mathscr{I}/\varepsilon$. The results obtained in the previous sections for $\mathscr{G}$ and $\mathscr{I}$ apply to $\mathscr{E}$ and $\mathscr{E}^2$ by simply replacing ε with unity in all the expressions (as was done in the examples). If this normalization is made in resolution shells, then the decline in structure factors with resolution is removed. Another form of normalization is

$$\mathscr{E}=\mathscr{G}/\langle\mathscr{G}^2\rangle^{1/2}=(2/\varepsilon m)^{1/2}\mathscr{G} \quad (38)$$

or

$$\mathscr{E}^2=\mathscr{I}/\langle\mathscr{I}\rangle=2\mathscr{I}/\varepsilon m. \quad (39)$$

This has the disadvantage that the normalization depends on m, but the advantage that

$$\langle\mathscr{E}^2\rangle=1 \quad (40)$$

independently of m. This may be useful when considering intensity distributions over the whole diffraction pattern rather than for particular values of m. The probability density for $\mathscr{E}$ defined by (38), for example, is

$$P_m(\mathscr{E})=[2/\Gamma(m/2)](m/2)^{m/2}\mathscr{E}^{m-1}\times\exp(-m\mathscr{E}^2/2). \quad (41)$$

Any normalization applied to fiber diffraction amplitudes will probably depend on the specific application of intensity statistics.

6. Discussion

Although Figs. 1 and 2 provide an indication of the accuracy of the normal approximations to the distributions of amplitudes and intensities, the accuracy can be more succinctly summarized by examining the normalized r.m.s. error, E_m, defined by

$$E_m^2=\int_0^\infty[P_m(x)-\hat{P}_m(x)]^2\,dx\times\left\{\int_0^\infty[P_m(x)]^2\,dx\right\}^{-1} \quad (42)$$

where x is either $\mathscr{G}$ or $\mathscr{I}$. These errors were calculated numerically for the different normal approximations to the probability density functions, and are shown in Fig. 6. They illustrate that, for $P(\mathscr{G})$, the error is modest for typical values of m. (The mean value of m is usually between about 4 and 8 for high-resolution studies - see below.) The errors are significantly lower for the approximation $\hat{P}'_m(\mathscr{G})$ than for $\hat{P}_m(\mathscr{G})$. Whether these errors are acceptable would depend on the specific application of the amplitude distributions. The errors in the normal approximation to $P_m(\mathscr{I})$ are rather large, suggesting that in applications involving intensities one would probably have to use either the exact expression or the Edgeworth series approximation, whichever is more convenient.

To further assess the significance of these errors, the values of m on typical fiber diffraction patterns must be considered. Since m depends on the position in reciprocal space, its overall effect can be estimated by considering its mean value over the diffraction pattern, $\langle m \rangle$, given by

$$\langle m \rangle = [1/(L+1)] \sum_{l=0}^{L} [1/(R_{\max} - R_{\min})] \times \int_{R_{\min}}^{R_{\max}} m(l, R)\, dR \qquad (43)$$

where $m(l, R)$ denotes the value of m on layer line l at reciprocal-space cylindrical radius R, $R_{\min}$ and $R_{\max}$ are the minimum and maximum values of R on layer line l between the minimum and maximum resolution limits of the diffraction data and L is the maximum layer-line number. The value of $\langle m \rangle$ depends on the molecular diameter and symmetry, the c repeat (c) and the resolution limits of the diffraction data (Stubbs, 1989). The dependence of $\langle m \rangle$ on maximum resolution is shown in Fig. 7(a) for two structures; a nucleic acid (Park, Arnott, Chandrasekaran, Millane & Campagnari 1987; diameter = 20 Å, $c = 32{\cdot}3$ Å, 10_1 helix symmetry and minimum resolution = 20 Å), and tobacco mosaic virus (TMV; Namba & Stubbs, 1985; diameter = 180 Å, $c = 69{\cdot}0$ Å, 49_3 helix symmetry and minimum resolution = 10 Å). These represent typical medium-sized and large molecules, respectively, studied by fiber diffraction, and Fig. 7(a) shows that $\langle m \rangle$ is typically between 4 and 8 for high-resolution studies. The corresponding errors, $E_{\langle m \rangle}$, for these two structures are shown in Fig. 7(b). These show that for high-resolution studies, the normal approximations to $P_m(G)$ are probably sufficiently accurate [particularly $\hat{P}'_m(\mathscr{G})$] for many purposes. As noted above, normal approximations to $P_m(\mathscr{I})$ are not as accurate and should be used with caution.

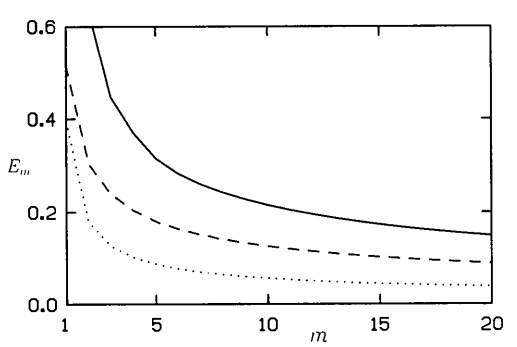

Fig. 6. Normalized r.m.s. error E_m for the normal approximations to the intensities $\hat{P}_m(\mathscr{I})$ (——), and the amplitudes $\hat{P}_m(\mathscr{G})$ (- - -) and $\hat{P}'_m(\mathscr{G})$ (· · ·) on a fiber diffraction pattern as a function of m.

7. Concluding remarks

A number of properties of the distributions of amplitudes and intensities on a fiber diffraction pattern have been derived. Probably the most significant result is that, for a particular m, the amplitudes and intensities are approximately normally distributed.

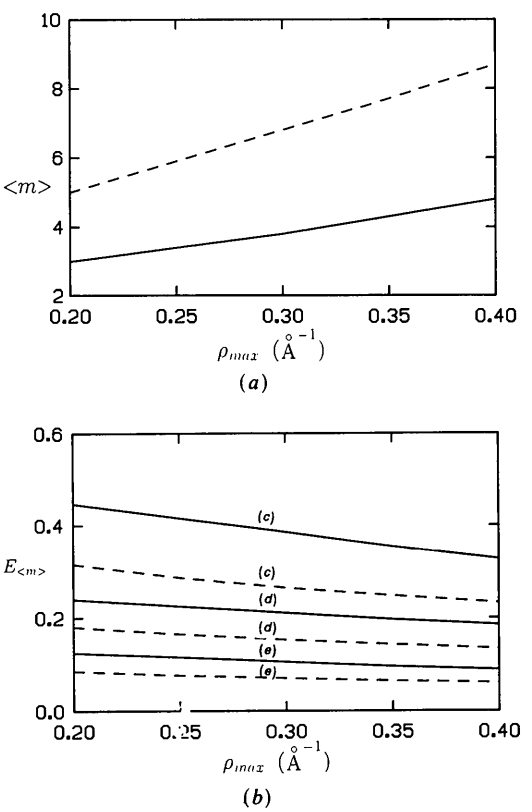

Fig. 7. (a) The mean value of m, $\langle m \rangle$, and (b) the r.m.s. error, $E_{\langle m \rangle}$, as a function of maximum resolution for the DNA (——) and TMV (- - -) structures (see text). The different curves in (b) refer to the normal approximations (c) $\hat{P}_m(\mathscr{I})$, (d) $\hat{P}_m(\mathscr{G})$, and (e) $\hat{P}'_m(\mathscr{G})$.

The general simplifications that result when dealing with normal distributions may be useful in applications of amplitude or intensity statistics. Since the utility of these approximations would probably result from their simplicity, the normal approximations (19), (32) and (34) may be more useful than the other more accurate, but more complex, approximations derived here. Although the Edgeworth series approximations are more complicated than the exact expressions, they may be useful in some circumstances because they take the form of a correction to a normal distribution. The examples presented indicate that the normal approximations may be sufficiently accurate (especially for the amplitudes) in practical applications. The other properties derived here may also be useful in specific applications.

The analysis presented here indicates some limitations in the application of intensity statistics in fiber diffraction. Since the distributions are approximately normal for typical values of m, the *shapes* of density or distribution curves are largely independent of m. Hence m probably cannot be reliably estimated from the shapes of the distributions (as can be done in traditional crystallography, for example, to distinguish between centrosymmetric and non-centrosymmetric crystals). This is illustrated vividly in Fig. 5, which shows that the cumulative distribution functions for different m would superimpose almost exactly by changing the origin for $\mathcal{G}$. (The central limit theorem is 'a great equalizer'.) Fig. 3 shows that the intensities would be more effective than the amplitudes if attempting to do this. It may be possible, however, to estimate variations in m by examining the variation of the mean, variance or central moments over the diffraction pattern. (Note that the variance of the amplitude is not suitable for this purpose, however, as it is largely independent of m.) This could be useful in resolving ambiguities in symmetry as discussed below. One would also have to consider limitations imposed by the small number of independent data on a fiber diffraction pattern that may make it difficult to obtain a statistically significant number of samples.

Intensity distributions in fiber diffraction have been used to estimate largest likely R factors, and other applications are possible. These include estimation of atomic coordinate errors, and determination of symmetry. Although helix symmetry can usually be determined from the distribution of meridional reflections, this is not always straightforward. Also, meridional reflections do not distinguish between integral and non-integral helices. Appropriate use of intensity statistics may help resolve some of these ambiguities by analyzing the variation of m over the diffraction pattern. Some of the results derived here may also be useful in optimizing procedures such as isomorphous replacement and difference Fourier synthesis in fiber diffraction. The use of intensity statistics to phase fiber diffraction data is probably limited (as in protein crystallography), although it is possible that they could be used to supplement conventional phasing techniques.

I am grateful to the US National Science Foundation for support (DMB-8606942) and Deb Zerth for word processing.

APPENDIX *A*
Normal approximation to $P(\mathcal{G})$

From (8)

$$\mathcal{G} = \left(\sum_{i=1}^{m/2} \mathcal{I}_i \right)^{1/2} \qquad (A.1)$$

and the random variable $\mathcal{G}$ can be considered a function of the $m/2$ random variables $\mathcal{I}_i$. Although it is assumed here that m is even, extension to odd m is straightforward. This function is expanded as a multidimensional Taylor series about the $(m/2)$-dimensional vector $\boldsymbol{\nu}$ of mean values of the $\mathcal{I}_i$ giving

$$\mathcal{G} = \mathcal{G}(\boldsymbol{\nu}) + \sum_{i=1}^{m/2} [\partial \mathcal{G}(\boldsymbol{\nu})/\partial \mathcal{I}_i](\mathcal{I}_i - \nu^{(i)}) + R \qquad (A.2)$$

where the $\nu^{(i)}$ are the components of $\boldsymbol{\nu}$. The remainder term R is given by

$$R = (1/2) \sum_{i=1}^{m/2} \sum_{j=1}^{m/2} [\partial^2 \mathcal{G}(\boldsymbol{\nu}_{ij})/\partial \mathcal{I}_i \partial \mathcal{I}_j] \times (\mathcal{I}_i - \nu^{(i)})(\mathcal{I}_j - \nu^{(j)}) \qquad (A.3)$$

where

$$\boldsymbol{\nu}_{ij} = (\nu^{(1)}, \nu^{(2)}, \ldots, \xi_i, \ldots, \zeta_j, \ldots, \nu^{(m/2)}), \qquad (A.4)$$

the ξ_i and ξ_j are the ith and jth components respectively of $\boldsymbol{\nu}_{ij}$, and

$$0 < \nu^{(i)} \gtrless \xi_i, \zeta_i \gtrless \mathcal{I}_i > 0, \qquad (A.5)$$

i.e. ξ_i and ζ_i are between $\nu^{(i)}$ and $\mathcal{I}_i$, and all the quantities are positive. Since each $\mathcal{I}_i$ has two degrees of freedom ($m = 2$), from (13) $\nu^{(i)} = \varepsilon$, and use of (*A*.1) to evaluate the partial derivatives in (*A*.2) gives

$$\mathcal{G} = (\varepsilon m/2)^{1/2} + (2\varepsilon m)^{-1/2} \sum_{i=1}^{m/2} (\mathcal{I}_i - \varepsilon) + R. \qquad (A.6)$$

The first term in (*A*.6) is a constant and the second is a sum of identically distributed zero-mean random variables. By the central limit theorem therefore, the first two terms are normally distributed as $m \to \infty$, with mean $(\varepsilon m/2)^{1/2}$ and variance $\varepsilon/4$. A similar evaluation of the remainder term gives

$$R = -(1/8) \sum_{i=1}^{m/2} \sum_{j=1}^{m/2} [\varepsilon(m - 2 + \xi_i + \zeta_j)/2]^{-3/2} \times (\mathcal{I}_i - \varepsilon)(\mathcal{I}_j - \varepsilon). \qquad (A.7)$$

Since, from (*A*.5), $\xi_i, \zeta_j > 0$,

$$|R| < (\sqrt{2}/4)[\varepsilon(m-2)]^{-3/2} \sum_{i=1}^{m/2} \sum_{j=1}^{m/2} (\mathcal{I}_i - \varepsilon)(\mathcal{I}_j - \varepsilon), \quad (A.8)$$

the last factor being the sum of $m^2/4$ zero-mean random variables. The remainder term is therefore normally distributed as $m \to \infty$ with zero mean and a variance that is $O(m^{-1})$. Hence, R in (*A*.6) is insignificant compared with the first two terms as $m \to \infty$, so that $\mathcal{G}$ is normally distributed in the limit, with the mean and variance given above. Analysis of the third central moment of the second term in (*A*.6) shows that its leading behavior is $(1/8)2^{-1/2}\varepsilon^{3/2}m^{-1/2}$ as $m \to \infty$.

APPENDIX *B*

Asymptotic behavior of $\varphi(y)$

The asymptotic behavior of the characteristic function $\varphi_m(y)$ for $\mathcal{G}$, given exactly by (31), as $m \to \infty$ is derived here. The asymptotic expansion for the parabolic cylinder function (Abramowitz & Stegun, 1972, equation 19.9.1) shows that

$$D_{-m}(x) = \frac{\sqrt{\pi}\exp[-(m-1/2)^{1/2}x]}{2^{m/2}\Gamma(m/2+1/2)} \times \left[1 - \frac{x^3}{24m^{1/2}} + O(m^{-1})\right] \quad m \to \infty. \quad (B.1)$$

Development of the exponential as an asymptotic series in m gives

$$D_{-m}(x) = \frac{\sqrt{\pi}\exp(-m^{1/2}x)}{2^{m/2}\Gamma(m/2+1/2)} \times \left[1 + \frac{x}{4}\left(1 - \frac{x^2}{6}\right)m^{-1/2} + O(m^{-1})\right] \quad m \to \infty. \quad (B.2)$$

Substitution of (*B*.2) into (31) gives

$$\varphi_m(y) = \exp[i(\varepsilon m/2)^{1/2}y]\exp(-\varepsilon y^2/8) \times \left[1 - \frac{i\varepsilon^{1/2}y}{4\sqrt{2}}\left(1 + \frac{\varepsilon y^2}{12}\right)m^{-1/2} + O(m^{-1})\right] \quad m \to \infty \quad (B.3)$$

which gives the first two terms in the asymptotic series for $\varphi(y)$. Comparison of (*B*.3) with the approximate characteristic function (37) shows that the latter is the first term in the asymptotic expansion. The coefficient of $m^{-1/2}$ in (*B*.3) is related to the third-order Hermite polynomial in the Edgeworth series (33) for $P_m(\mathcal{G})$.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.

BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410-445.

COCHRAN, W., CRICK, F. H. C. & VAND, V. (1952). *Acta Cryst.* **5**, 581-586.

CRAMER, H. (1970). *Random Variables and Probability Distributions*, 3rd ed. Cambridge Univ. Press.

FRANKLIN, R. E. & KLUG, A. (1955). *Acta Cryst.* **8**, 777-780.

GIACOVAZZO, C. (1980). *Direct Methods in Crystallography.* London: Academic Press.

GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Table of Integrals, Series and Products.* New York: Academic Press.

HAUPTMAN, H. & KARLE, J. (1953). *The Solution of the Phase Problem*: I. *The Centrosymmetric Crystal. Am. Crystallogr. Assoc. Monogr.* No. 3. Pittsburgh: Polycrystal Book Service.

KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.

KLUG, A., CRICK, F. H. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* **11**, 199-213.

LUZATTI, P. V. (1952). *Acta Cryst.* **5**, 802-810.

MAKOWSKI, L. (1982). *J. Appl. Cryst.* **15**, 546-557.

MILLANE, R. P. (1988). *Computing in Crystallography* 4: *Techniques and New Technologies*, edited by N. W. ISAACS & M. R. TAYLOR, pp. 169-186. IUCr/Oxford Univ. Press.

MILLANE, R. P. (1989*a*). *Acta Cryst.* A**45**, 258-260.

MILLANE, R. P. (1989*b*). *Acta Cryst.* A**45**, 573-576.

MILLANE, R. P. (1990). *Acta Cryst.* A**46**, 68-72.

NAMBA, K. & STUBBS, G. (1985). *Acta Cryst.* A**41**, 252-262.

NAMBA, K. & STUBBS, G. (1987). *Acta Cryst.* A**43**, 533-539.

PARK, H. S., ARNOTT, S., CHANDRASEKARAN, R., MILLANE, R. P. & CAMPAGNARI, F. (1987). *J. Mol. Biol.* **197**, 513-523.

STUBBS, G. (1989). *Acta Cryst.* A**45**, 254-258.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.

WILSON, A. J. C. (1950). *Acta Cryst.* **3**, 397-399.

YEATES, T. O. (1988). *Acta Cryst.* A**44**, 142-144.