Acta Cryst. (1999). A**55**, 322–331

The probability distribution function of structure factors with non-integral indices. III. The joint probability distribution in the $P\bar{1}$ case

Carmelo Giacovazzo,[a]* Dritan Siliqi,[a,b] Angela Altomare,[a] Giovanni Luca Cascarano,[a] Rosanna Rizzi[a] and Riccardo Spagna[c]

[a]*IRMEC, c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy,* [b]*Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania, and* [c]*Istituto di Strutturistica Chimica 'G. Giacomello', CNR, CP 10, Monterotondo Stazione, 00016 Roma, Italy. E-mail: c.giacovazzo@area.ba.cnr.it*

(*Received 24 April 1998; accepted 23 July 1998*)

Abstract

The joint probability distribution function method has been developed in $P\bar{1}$ for reflections with rational indices. The positional atomic parameters are considered to be the primitive random variables, uniformly distributed in the interval (0, 1), while the reflection indices are kept fixed. Owing to the rationality of the indices, distributions like $P(F_{\mathbf{p}_1}, F_{\mathbf{p}_2})$ are found to be useful for phasing purposes, where $\mathbf{p}_1$ and $\mathbf{p}_2$ are any pair of vectorial indices. A variety of conditional distributions like $P(|F_{\mathbf{p}_1}|\,||F_{\mathbf{p}_2}|)$, $P(|F_{\mathbf{p}_1}|\,|F_{\mathbf{p}_2})$, $P(\varphi_{\mathbf{p}_1}|\,|F_{\mathbf{p}_1}|, F_{\mathbf{p}_2})$ are derived, which are able to estimate the modulus and phase of $F_{\mathbf{p}_1}$ given the modulus and/or phase of $F_{\mathbf{p}_2}$. The method has been generalized to handle the joint probability distribution of any set of structure factors, *i.e.* the distributions $P(F_1, F_2, \ldots, F_{n+1})$, $P(|F_1|\,|F_2, \ldots, F_{n+1})$ and $P(\varphi_1|\,|F|_1, F_2, \ldots, F_{n+1})$ have been obtained. Some practical tests prove the efficiency of the method.

1. Symbols and notation

N: number of atoms in the unit cell
f_j: scattering factor of the jth atom (thermal factor included)
$\mathbf{h}$: three-dimensional index with integral components (h, k, l)
$\mathbf{p}$, $\mathbf{q}$: three-dimensional indices with rational components (p_1, p_2, p_3), (q_1, q_2, q_3), respectively
$p_s = p_1 + p_2 + p_3$
$q_s = q_1 + q_2 + q_3$
φ: phase of the structure factor

$$\Sigma_1(\mathbf{p}),\ \Sigma_1(\mathbf{q}) = \sum_{j=1}^{N} f_j$$

$$\Sigma_2(\mathbf{p}),\ \Sigma_2(\mathbf{q}) = \sum_{j=1}^{N} f_j^2$$

$$\Sigma_{11}(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^{N} f_j(\mathbf{p}) f_j(\mathbf{q}).$$

Papers by Giacovazzo & Siliqi (1998) and Giacovazzo *et al.* (1999) will be referred to as paper I and paper II, respectively.

2. Introduction

Wilson's statistics (Wilson, 1942) may be considered the first step for the development of the direct-methods procedures. They allow diffraction data to be put on the absolute scale and provide an estimate of the overall Debye–Waller factor. If the atomic positions are assumed to be the primitive random variables while the reflections indices are kept fixed, Wilson's statistical method is able to predict the most probable value of the structure-factor modulus, but is unable to estimate its phase. The joint probability distribution method of the structure factors (Hauptman & Karle, 1953) filled the gap: estimates of both moduli and phases became available, provided the indices of the reflections involved in the distribution give rise to structure invariants or seminvariants. The practical procedures proved to be quite efficient: today, direct methods are considered the most efficient tool for solving up to medium-size crystal structures.

In papers I and II of this series, the probability distribution $P(E_{\mathbf{p}})$ has been derived in $P1$ and $P\bar{1}$, respectively, for rational values of p_1, p_2, p_3. For integral values of the indices, the distribution coincides with Wilson's distribution, but it may strongly differ if the indices are (or are close to) half-integers. The deviations are quite remarkable if the integral parts of the indices are small, and increase with the size of the structure. Unlike in Wilson statistics, in favourable circumstances both the moduli and the phases can be accurately estimated. Papers I and II show that Wilson's statistics are part of a larger family, the statistics of the structure factors with rational indices.

It is now natural to consider papers I and II of this series as the first step for studying a wider field, *i.e.* the development of direct methods applied to reflections with rational indices. To this end, the joint probability

ISSN 0108-7673 © 1999

distribution of structure factors must be developed, by allowing the indices to vary over the set of rational numbers. This paper starts such a study in $P\bar{1}$. We will see that the indices of the reflections involved in the joint probability distribution functions do not necessarily give rise to structure invariants or seminvariants. Indeed, phase as well as modulus estimates can be obtained by considering pairs of reflections: this additional degree of freedom is allowed by the rationality of the indices and by our assumptions on the primitive random variables. Joint probability distribution functions of structure factors giving rise to structure invariants can also be useful, but they will be examined in a future paper.

We will first derive here the joint probability distribution $P(F_{\mathbf{p}}, F_{\mathbf{q}})$, where $\mathbf{p}$ and $\mathbf{q}$ can be any pair of vectors with rational components, the case of integral component included. We will show that, in favourable circumstances, $F_{\mathbf{p}}$ and $F_{\mathbf{q}}$ are statistically correlated with each other (§3). We will then derive a family of related distributions, the most promising of which (§6) are able to estimate the phase and modulus of a structure factor when other phases and moduli are *a priori* known. Some initial experimental tests are also described and prove the efficiency of our method.

The possible applications of the probabilistic relationships established in this paper are not completely defined. The reader will find a wide report in the *Introduction* of paper I.

3. The distribution $P(F_{\mathbf{p}}, F_{\mathbf{q}})$

Let us suppose that the indices $\mathbf{p}$ and $\mathbf{q}$ are kept fixed and that the variables x_j, y_j, z_j, $j = 1, \ldots, N/2$, are independently and uniformly distributed in the interval $(0, 1)$. In order to satisfy the latter condition for all the atoms in the unit cell [*i.e.* x_j, y_j, z_j, $j = 1, \ldots, N$, distributed in the interval $(0, 1)$], the $N/2$ symmetry-equivalent atoms are generated by applying the inversion centre at $(1/2, 1/2, 1/2)$. Then (see paper II),

$$F_{\mathbf{p}} = A_{\mathbf{p}} + iB_{\mathbf{p}} \qquad F_{\mathbf{q}} = A_{\mathbf{q}} + iB_{\mathbf{q}},$$

where

$$A_{\mathbf{p}} = \cos(\pi p_s)A_{0\mathbf{p}}, \qquad B_{\mathbf{p}} = \sin(\pi p_s)A_{0\mathbf{p}}, \qquad (1)$$

$$A_{\mathbf{q}} = \cos(\pi q_s)A_{0\mathbf{q}}, \qquad B_{\mathbf{q}} = \sin(\pi q_s)A_{0\mathbf{q}}, \qquad (2)$$

$$A_{0\mathbf{p}} = 2\sum_{j=1}^{N/2} f_j \cos[(\pi p_s) - 2\pi \mathbf{p}\cdot\mathbf{r}_j]. \qquad (3)$$

The phases of $F_{\mathbf{p}}$ and $F_{\mathbf{q}}$ are given by

$$\varphi_{\mathbf{p}} = \tan^{-1}\{[\sin(\pi p_s)A_{0\mathbf{p}}]/[\cos(\pi p_s)A_{0\mathbf{p}}]\}, \qquad (4)$$

$$\varphi_{\mathbf{q}} = \tan^{-1}\{[\sin(\pi q_s)A_{0\mathbf{q}}]/[\cos(\pi q_s)A_{0\mathbf{q}}]\}, \qquad (5)$$

respectively. If the distribution $P(A_{0\mathbf{p}}, A_{0\mathbf{q}})$ is known, the other distributions of interest easily follow. Indeed:

(*a*) The distribution $P(A_{\mathbf{p}}, A_{\mathbf{q}})$ is obtained through the change of variables given by the relations (1) and (2).

(*b*) Since $B_{\mathbf{p}}$ and $B_{\mathbf{q}}$ are algebraically related to $A_{\mathbf{p}}$ and $A_{\mathbf{q}}$ *via*

$$B_{\mathbf{p}} = A_{\mathbf{p}}\tan(\pi p_s), \qquad B_{\mathbf{q}} = A_{\mathbf{q}}\tan(\pi q_s),$$

the distribution $P(A_{\mathbf{p}}, B_{\mathbf{p}}, A_{\mathbf{q}}, B_{\mathbf{q}})$ is simply

$$P(A_{\mathbf{p}}, A_{\mathbf{q}})\,\delta[B_{\mathbf{p}} - A_{\mathbf{p}}\tan(\pi p_s)]\,\delta[B_{\mathbf{q}} - A_{\mathbf{q}}\tan(\pi q_s)],$$

where δ is the Dirac delta function. Thus, $P(A_{\mathbf{p}}, B_{\mathbf{p}}, A_{\mathbf{q}}, B_{\mathbf{q}})$ is completely known when $P(A_{\mathbf{p}}, A_{\mathbf{q}})$ is known.

(*c*) Since

$$|F_{\mathbf{p}}|^2 = A_{\mathbf{p}}^2 + B_{\mathbf{p}}^2 = A_{0\mathbf{p}}^2, \quad |F_{\mathbf{q}}|^2 = A_{\mathbf{q}}^2 + B_{\mathbf{q}}^2 = A_{0\mathbf{q}}^2, \quad (6)$$

the distribution $P(|F_{\mathbf{p}}|, |F_{\mathbf{q}}|)$ is fixed once $P(A_{0\mathbf{p}}, A_{0\mathbf{q}})$ is available. Also the phases $\varphi_{\mathbf{p}}$ and $\varphi_{\mathbf{q}}$ are algebraically fixed by (4) and (5) when $A_{0\mathbf{p}}$ and $A_{0\mathbf{q}}$ are known. Indeed,

$$F_{\mathbf{p}} = |F_{\mathbf{p}}|\exp(i\varphi_{\mathbf{p}}) = A_{0\mathbf{p}}\exp(i\pi p_s),$$

$$F_{\mathbf{q}} = A_{0\mathbf{q}}\exp(i\pi q_s),$$

from which

$$A_{0\mathbf{p}} = |F_{\mathbf{p}}|\exp[i(\varphi_{\mathbf{p}} - \pi p_s)] = |F_{\mathbf{p}}|\cos(\varphi_{\mathbf{p}} - \pi p_s),$$

$$A_{0\mathbf{q}} = |F_{\mathbf{q}}|\cos(\varphi_{\mathbf{q}} - \pi q_s)$$

are obtained. As a consequence, $\mathbf{P}(\varphi_{\mathbf{p}}, \varphi_{\mathbf{q}}\,|\,|F_{\mathbf{p}}|, |F_{\mathbf{q}}|)$ is also known if $P(A_{0\mathbf{p}}, A_{0\mathbf{q}})$ is known.

The characteristic function $C(u_1, u_2)$ of the distribution $P(A_{0\mathbf{p}}, A_{0\mathbf{q}})$ may be written in terms of the cumulants K_{ij} of the distribution. If only terms up to the second order are considered, we have

$$\begin{aligned} C(u_1, u_2) = {} & \exp[i(K_{10}u_1 + K_{01}u_2) \\ & - (K_{20}u_1^2 + K_{02}u_2^2 + 2K_{11}u_1u_2)/2], \end{aligned}$$

where

$$K_{10} = m_{10} = \Sigma_1(\mathbf{p})c_{p_1/2}c_{p_2/2}c_{p_3/2},$$

$$K_{01} = m_{01} = \Sigma_1(\mathbf{q})c_{q_1/2}c_{q_2/2}c_{q_3/2},$$

$$K_{20} = m_{20} - m_{10}^2 = \Sigma_2(\mathbf{p})(1 + c_{p_1}c_{p_2}c_{p_3} - 2c_{p_1/2}^2c_{p_2/2}^2c_{p_3/2}^2),$$

$$K_{02} = m_{02} - m_{01}^2 = \Sigma_2(\mathbf{q})(1 + c_{q_1}c_{q_2}c_{q_3} - 2c_{q_1/2}^2c_{q_2/2}^2c_{q_3/2}^2),$$

$$\begin{aligned} K_{11} &= m_{11} - m_{10}m_{01} \\ &= \Sigma_{12}(\mathbf{p}, \mathbf{q})[-2c_{p_1/2}c_{p_2/2}c_{p_3/2}c_{q_1/2}c_{q_2/2}c_{q_3/2} \\ &\quad + c_{(p_1+q_1)/2}c_{(p_2+q_2)/2}c_{(p_3+q_3)/2} \\ &\quad + c_{(p_1-q_1)/2}c_{(p_2-q_2)/2}c_{(p_3-q_3)/2}], \end{aligned}$$

$$c_{p_i} = \sin(2\pi p_i)/(2\pi p_i),$$

$$c_{q_i} = \sin(2\pi q_i)/(2\pi q_i).$$

m_{ij} are the joint moments of the distribution. The reader will find in Appendix *A* of paper II the technique for the derivation of $m_{10}, m_{01}, m_{20}, m_{02}$; the mixed moment m_{11} is derived in Appendix *A* of this paper. By repeated application of (35), we obtain

$$\begin{aligned} P(A_{0\mathbf{p}}, A_{0\mathbf{q}}) &\simeq (2\pi)^{-2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(u_1, u_2) \\ &\quad \times \exp(-iu_1 A_{0\mathbf{p}} - iu_2 A_{0\mathbf{q}})\, \mathrm{d}u_1\, \mathrm{d}u_2 \\ &= (2\pi)^{-1} \Delta^{-1/2} \exp\{-(2\Delta)^{-1} \\ &\quad \times [K_{02}(A_{0\mathbf{p}} - K_{10})^2 + K_{20}(A_{0\mathbf{q}} - K_{01})^2 \\ &\quad - 2K_{11}(A_{0\mathbf{p}} - K_{10})(A_{0\mathbf{q}} - K_{01})]\}, \qquad (7) \end{aligned}$$

where

$$\Delta = \begin{vmatrix} K_{20} & K_{11} \\ K_{11} & K_{02} \end{vmatrix} = K_{20}K_{02} - K_{11}^2 \geq 0.$$

Equation (7) is the first basic result of this paper. We note:

(*a*) Equation (7) is a two-dimensional normal distribution, with expected marginal moments $\langle A_{0\mathbf{p}} \rangle = K_{10}$ and $\langle A_{0\mathbf{q}} \rangle = K_{01}$.

(*b*) The cumulant K_{11} is in general different from zero. Thus the knowledge of $A_{0\mathbf{q}}$ may provide information on $A_{0\mathbf{p}}$ and *vice versa*.

(*c*) K_{11} regulates the correlation between $A_{0\mathbf{p}}$ and $A_{0\mathbf{q}}$. The K_{11} value depends on the resolution of the reflections **p** and **q** through the term $\Sigma_{11}(\mathbf{p}, \mathbf{q})$, and on the type of the indices (through the *c* coefficients). The three contributions in the K_{11} expression may be so interpreted: $c_{(p_1-q_1)/2}c_{(p_2-q_2)/2}c_{(p_3-q_3)/2}$ is due to the simultaneous knowledge of $A_{0\mathbf{p}}$ and $A_{0\mathbf{q}}$ (and depends on the relative position of **p** and **q** in the reciprocal space), $c_{(p_1+q_1)/2}c_{(p_2+q_2)/2}c_{(p_3+q_3)/2}$ is due to the consequent knowledge of the pair $A_{0\mathbf{p}}$ and $A_{0(-\mathbf{q})}$ (and depends on the relative position of **p** and $-\mathbf{q}$ in the reciprocal space), the term $c_{p_1/2}c_{p_2/2}c_{p_3/2}c_{q_1/2}c_{q_2/2}c_{q_3/2}$ is the amount of correlation generated by prior information on F_{000}. Indeed, this latter information produces the estimates $\langle A_{0\mathbf{p}} \rangle = K_{10}$ and $\langle A_{0\mathbf{q}} \rangle = K_{01}$ and, consequently, their combination in K_{11}; this third contribution depends on the absolute positions of **p** and **q** in reciprocal space.

If both **p** and **q** have integral components ($\mathbf{p} = \mathbf{h}$ and $\mathbf{q} = \mathbf{k}$), then

$$c_{p_i} = c_{q_i} = c_{p_i/2} = c_{q_i/2} = c_{(p_i+q_i)/2} = c_{(p_i-q_i)/2} = 0$$

for $i = 1, 2, 3$.

Accordingly,

$$K_{10} = K_{01} = 0 \qquad K_{20} = \Sigma_2(\mathbf{h}),$$
$$K_{02} = \Sigma_2(\mathbf{k}), \qquad K_{11} = 0,$$

and (7) reduces to the product of two uncorrelated Wilson's distributions,

$$\begin{aligned} P(A_{0\mathbf{h}}, A_{0\mathbf{k}}) &= (2\pi)^{-1}[\Sigma_2(\mathbf{h})\Sigma_2(\mathbf{k})]^{-1/2} \\ &\quad \times \exp[-A_{0\mathbf{h}}^2/2\Sigma_2(\mathbf{h}) - A_{0\mathbf{k}}^2/2\Sigma_2(\mathbf{k})]. \end{aligned}$$

K_{11} is different from zero when **p** has integral and **q** has non-integral components, or *vice versa*. A particular case of great interest (the *canonical case*) occurs when one of **p** and **q** has half-integral indices and the other has integral indices; this case will be treated in §7.

When **p** is very close to **q**, the correlation between $A_{0\mathbf{p}}$ and $A_{0\mathbf{q}}$ increases and

$$K_{11} = \langle A_{0\mathbf{p}} A_{0\mathbf{q}} \rangle \cong \langle A_{0\mathbf{p}}^2 \rangle = K_{20}.$$

In agreement with this result, $(\mathbf{p} + \mathbf{q}) \cong 2\mathbf{p}$, $(\mathbf{p} - \mathbf{q}) \cong 0$ and our K_{11} expression reduces to K_{20}. The distribution (7) has to be singular when $\mathbf{p} \equiv \mathbf{q}$; accordingly, the condition $\Delta = 0$ will make (7) singular.

4. The distributions $P(A_\mathbf{p}, A_\mathbf{q})$ and $P(|F_\mathbf{p}|||F_\mathbf{q}|)$

Several distributions can be inferred from (7): as examples, we shortly describe here how to obtain $P(A_\mathbf{p}, A_\mathbf{q})$ and $P(|F_\mathbf{p}|, |F_\mathbf{q}|)$.

In accordance with (1) and (2), we can write (provided p_s and q_s are not half-integers)

$$\begin{aligned} P(A_\mathbf{p}, A_\mathbf{q}) &= (2\pi)^{-1}\Delta^{-1/2}[\cos(\pi p_s)\cos(\pi q_s)]^{-1} \\ &\quad \times \exp\big(-(2\Delta)^{-1}\{K_{02}[A_\mathbf{p}/\cos(\pi p_s) - K_{10}]^2 \\ &\quad + K_{20}[A_\mathbf{q}/\cos(\pi q_s) - K_{01}]^2 \\ &\quad - 2K_{11}[A_\mathbf{p}/\cos(\pi p_s) - K_{10}] \\ &\quad \times [A_\mathbf{q}/\cos(\pi q_s) - K_{01}]\}\big). \end{aligned}$$

If p_s and q_s are half-integers, $A_\mathbf{p} \equiv 0$ and $A_\mathbf{q} \equiv 0$: the distribution $P(A_\mathbf{p}, A_\mathbf{q})$ may then also be calculated in this case.

Because of (6),

$$\begin{aligned} P(|F_\mathbf{p}|, |F_\mathbf{q}|) &= P(A_{0\mathbf{p}}, A_{0\mathbf{q}}) + P(-A_{0\mathbf{p}}, A_{0\mathbf{q}}) \\ &\quad + P(-A_{0\mathbf{p}}, -A_{0\mathbf{q}}) + P(A_{0\mathbf{p}}, -A_{0\mathbf{q}}). \end{aligned}$$

Then,

$$P(F_\mathbf{p}||F_\mathbf{q}|) = P(|F_\mathbf{p}|, |F_\mathbf{q}|) \Big/ \int_0^\infty P(|F_\mathbf{p}|, |F_\mathbf{q}|)\, \mathrm{d}|F_\mathrm{p}|. \qquad (8)$$

The integral in the denominator of (8) can be obtained by repeated application of the relation (37). We obtain

$$\begin{aligned} P(F_\mathbf{p}||F_\mathbf{q}|) &\cong (2/\pi)^{1/2}(K_{02}/\Delta)^{1/2} \\ &\quad \times \exp[-K_{02}|F_\mathbf{p}|^2/(2\Delta)]T/S, \qquad (9) \end{aligned}$$

where

$$T = \exp\{-[-K_{11}|F_{\mathbf{p}}F_{\mathbf{q}}| - (K_{02}K_{10} - K_{01}K_{11})|F_{\mathbf{p}}| - (K_{20}K_{01} - K_{10}K_{11})|F_{\mathbf{q}}|]/\Delta\} + \exp\{-[-K_{11}|F_{\mathbf{p}}F_{\mathbf{q}}| + (K_{02}K_{10} - K_{01}K_{11})|F_{\mathbf{p}}| + (K_{20}K_{01} - K_{10}K_{11})|F_{\mathbf{q}}|]/\Delta\} + \exp\{-[+K_{11}|F_{\mathbf{p}}F_{\mathbf{q}}| - (K_{02}K_{10} - K_{01}K_{11})|F_{\mathbf{p}}| + (K_{20}K_{01} - K_{10}K_{11})|F_{\mathbf{q}}|]/\Delta\} + \exp\{-[+K_{11}|F_{\mathbf{p}}F_{\mathbf{q}}| + (K_{02}K_{10} - K_{01}K_{11})|F_{\mathbf{p}}| - (K_{20}K_{01} - K_{10}K_{11})|F_{\mathbf{q}}|]/\Delta\},$$

$$S = 2\exp\{\Delta^{-1}[(K_{20}K_{01} - K_{10}K_{11})|F_{\mathbf{q}}| + (-K_{11}|F_{\mathbf{q}}| - K_{02}K_{10} + K_{01}K_{11})^2/(2K_{02})]\} + 2\exp\{\Delta^{-1}[(K_{20}K_{01} - K_{10}K_{11})|F_{\mathbf{q}}| - (-K_{11}|F_{\mathbf{q}}| + K_{02}K_{10} - K_{01}K_{11})^2/(2K_{02})]\}.$$

If $K_{11} = 0$ then the two diffraction moduli are uncorrelated and

$$P(|F_{\mathbf{p}}|||F_{\mathbf{q}}|) = [2/(\pi K_{20})]^{1/2}\exp[-K_{10}^2/(2K_{20})] \times\exp[-|F_{\mathbf{p}}|^2/(2K_{20})]\cosh(|F_{\mathbf{p}}|K_{10}/K_{20}),$$

which coincides with the $P(|F_{\mathbf{p}}|)$ distribution given in paper II [see equation (II.9)].

If $K_{11} \neq 0$, $|F_{\mathbf{p}}|$ may be estimated even if $|F_{\mathbf{q}}|$ is known and $\varphi_{\mathbf{q}}$ is unknown. This is obvious when **p** and **q** are very close to each other (*e.g.* $q_i = p_i \pm 0.1$ for $i = 1, 2, 3$), but less obvious in other cases.

5. The distribution $P(|F_{\mathbf{p}}||F_{\mathbf{q}})$, $P(\varphi_{\mathbf{p}}||F_{\mathbf{p}}|, F_{\mathbf{q}})$

Let us suppose that both the modulus and the sign of $A_{0\mathbf{q}}$ are *a priori* known. Then the conditional distribution $P(A_{0\mathbf{p}}|A_{0\mathbf{q}})$ may by calculated from (7). We obtain

$$P(A_{0\mathbf{p}}|A_{0\mathbf{q}}) \simeq (2\pi v_{0\mathbf{p}})^{-1/2}\exp[-(A_{0\mathbf{p}} - m_{0\mathbf{p}})^2/(2v_{0\mathbf{p}})], \quad (10)$$

where

$$v_{0\mathbf{p}} = \Delta/K_{02} \quad (11)$$

is the variance of the distribution (10), and

$$m_{0\mathbf{p}} = \langle A_{0\mathbf{p}}|A_{0\mathbf{q}}\rangle = K_{10} + K_{11}(A_{0\mathbf{p}} - K_{01})/K_{02} \quad (12)$$

is the expected value of $A_{0\mathbf{p}}$. We observe:

(i) The distribution (10) is a Gaussian, and reduces to (II.5) [*e.g.* to $P(A_{0\mathbf{p}})$], whether in absence of information on $A_{0\mathbf{q}}$ or when $K_{11} = 0$. Accordingly, if **p** has integral components, only reflections with non-integral indices will contribute to the estimation of $A_{0\mathbf{p}}$. A less obvious statement is the following: if the components of **p** are non-integral indices, all $A_{0\mathbf{q}}$ are correlated with $A_{0\mathbf{p}}$. To show that the statement is true, let us consider the case in which $\mathbf{p} \equiv (2.5, 1.5, 3.5)$. For any reflection **q** having half-integral indices, both $(p_i + q_i)$ and $(p_i - q_i)$, $i = 1, 2, 3$, are integers and consequently

$$c_{(p_i+q_i)/2} = c_{(p_i-q_i)/2} \equiv 0$$

for $i = 1, 2, 3$. However, $K_{11} \neq 0$ because

$$c_{p_1/2}c_{p_2/2}c_{p_3/2}c_{q_1/2}c_{q_2/2}c_{q_3/2} \neq 0.$$

Consider now $p \equiv (2.4, 1.6, 3.7)$. No reflection **q** may be found for which both $(p_i + q_i)$ and $(p_i - q_i)$, for $i = 1, 2, 3$, are integers, so that $K_{11} \neq 0$.

(ii) The expected value of $A_{0\mathbf{p}}$ is equal to K_{10} if $A_{0\mathbf{q}}$ is unknown. Prior knowledge of $A_{0\mathbf{q}}$ modifies the above expectation by the additional contribution $[K_{11}(A_{0\mathbf{q}} - K_{01})/K_{02}]$, which depends on the difference between $A_{0\mathbf{q}}$ and K_{01}. Reflections for which $A_{0\mathbf{q}}$ is very different from K_{01} will strongly contribute to establish the expected value of $A_{0\mathbf{p}}$. Since $K_{01} = 0$ for conventional reflections, their influence on the estimation of $F_{\mathbf{p}}$ will increase with $|F_{\mathbf{q}}|$.

(iii) The variance associated with the estimate of $A_{0\mathbf{p}}$ is $v_{0\mathbf{p}} = (K_{20} - K_{11}^2/K_{02})$. Since K_{20} is the variance of the estimate when $A_{0\mathbf{q}}$ is unknown, we obtain $v_{0\mathbf{p}} \leq K_{20}$: in simpler words, prior knowledge of $A_{0\mathbf{q}}$ improves the reliability of the estimate of $A_{0\mathbf{p}}$. The variance will remarkably decrease when K_{11} is large, which mostly occurs when $+\mathbf{q}$ or $-\mathbf{q}$ is very close to **p**.

Since $|F_{\mathbf{p}}| = |A_{0\mathbf{q}}|$, the distribution (10) may be used to calculate the distribution

$$P(|F_{\mathbf{p}}||A_{0\mathbf{q}}) \simeq [2/(\pi v_{0\mathbf{p}})]^{1/2}\exp[-(|F_{\mathbf{p}}|^2 + m_{0\mathbf{p}}^2)] \times \cosh(m_{0\mathbf{p}}|F_{\mathbf{p}}|/v_{0\mathbf{p}}), \quad (13)$$

from which [see (39)]

$$\langle|F_{\mathbf{p}}||A_{0\mathbf{q}}\rangle = m_{0\mathbf{p}}\Phi[m_{0\mathbf{p}}/(2v_{0\mathbf{p}})^{1/2}] + (2v_{0\mathbf{p}}/\pi)^{1/2}\exp[-m_{0\mathbf{p}}^2/2v_{0\mathbf{p}}]. \quad (14)$$

In practice, $F_{\mathbf{q}}$ is more directly accessible than $A_{0\mathbf{q}}$, therefore the distribution $P(|F_{\mathbf{p}}||F_{\mathbf{q}})$ may be more useful than (13). Owing to the identity

$$F_{\mathbf{q}} = A_{0\mathbf{q}}\exp(i\pi q_s),$$

we can rewrite (13) in the form

$$P(|F_{\mathbf{p}}||F_{\mathbf{q}}) \simeq [2/(\pi v_{0\mathbf{p}})]^{1/2}\exp[-(|F_{\mathbf{p}}|^2 + m_{0\mathbf{p}}^2)/(2v_{0\mathbf{p}})] \times \cosh(m_{0\mathbf{p}}|F_{\mathbf{p}}|/v_{0\mathbf{p}}), \quad (15)$$

where

$$m_{0\mathbf{p}} = K_{10} + K_{11}\{|F_{\mathbf{q}}|\exp[i(\varphi_{\mathbf{q}} - \pi q_s)] - K_{01}\}/K_{02} = K_{10} + K_{11}[|F_{\mathbf{q}}|\cos(\varphi_{\mathbf{q}} - \pi q_s) - K_{01}]/K_{02} \quad (16)$$

and

$$v_{0\mathbf{p}} = \Delta/K_{02}. \quad (17)$$

Equation (15) gives the probability distribution of a structure-factor modulus when another structure factor is known in modulus and phase. It allows the available information to be extrapolated beyond the experimental data, *e.g.* $F_{\mathbf{q}}$ may be a phased standard reflection and $F_{\mathbf{p}}$ a rational index reflection, or, *vice versa*, $F_{\mathbf{q}}$ may be a rational index reflection and $F_{\mathbf{p}}$ a standard index reflection at resolution higher than the experimental one.

On applying (39) to (15), we obtain

$$\langle|F_{\mathbf{p}}||F_{\mathbf{q}}\rangle = m_{0\mathbf{p}}\Phi[m_{0\mathbf{p}}/(2v_{0\mathbf{p}})^{1/2}] + (2v_{0\mathbf{p}}/\pi)^{1/2}\exp[-m_{0\mathbf{p}}^2/(2v_{0\mathbf{p}})], \quad (18)$$

with $m_{0\mathbf{p}}$ and $v_{0\mathbf{p}}$ given by (16) and (17).

On applying (40) to (15), the conditional expected value of $|F_{\mathbf{p}}|^2$ may be obtained,

$$\langle|F_{\mathbf{p}}|^2|F_{\mathbf{q}}\rangle = m_{0\mathbf{p}}^2 + v_{0\mathbf{p}}. \quad (19)$$

We see that the conditional expected value of $|F_{\mathbf{p}}|^2$ is particularly simple: it equals $m_{0\mathbf{p}}^2$ only if $v_{0\mathbf{p}}$ vanishes.

Let us now derive from (10) the distribution $P(\varphi_{\mathbf{p}}||F_{\mathbf{p}}|^2, F_{\mathbf{q}})$. Let $s_{0\mathbf{p}}$ be the sign of $A_{0\mathbf{p}}$. If we indicate by $P(s_{0\mathbf{p}} = 1)$ the probability that $s_{0\mathbf{p}}$ is equal to 1, we have

$$\begin{aligned} P(s_{0\mathbf{p}} = -1||A_{0\mathbf{p}}|, A_{0\mathbf{q}})/P(s_{0\mathbf{p}} = +1||A_{0\mathbf{p}}|, A_{0\mathbf{q}}) &\simeq \exp[-(|A_{0\mathbf{p}}| - m_{0\mathbf{p}})^2/(2v_{0\mathbf{p}})] \\ &\quad \times \{\exp[-(|A_{0\mathbf{p}}| - m_{0\mathbf{p}})^2/(2v_{0\mathbf{p}})]\}^{-1} \\ &\simeq \exp(-2|A_{0\mathbf{p}}|m_{0\mathbf{p}}/v_{0\mathbf{p}}). \end{aligned}$$

Since

$$\begin{aligned} P(s_{0\mathbf{p}} = +1||A_{0\mathbf{p}}|, A_{0\mathbf{q}}) &= \{1 + P(s_{0\mathbf{p}} = -1||A_{0\mathbf{p}}|, A_{0\mathbf{q}}) \\ &\quad \times [P(s_{0\mathbf{p}} = +1||A_{0\mathbf{p}}|, A_{0\mathbf{q}})]^{-1}\}^{-1}, \end{aligned}$$

we have

$$P(s_{0\mathbf{p}} = +1||A_{0\mathbf{p}}|, A_{0\mathbf{q}}) = 0.5 + 0.5\tanh(|A_{0\mathbf{p}}|m_{0\mathbf{p}}/v_{0\mathbf{p}}). \quad (20)$$

The relation (20) is the basis for a more useful relationship involving F instead of A. Since (*a*) the probability that $s_{0\mathbf{p}}$ is positive is exactly equal to the probability that $\varphi_{\mathbf{p}} = (\pi p_s)$, (*b*) $|A_{0\mathbf{p}}| = |F_{\mathbf{p}}|$, (*c*) $F_{\mathbf{q}} = A_{0\mathbf{q}}\exp(i\pi q_s)$ from which $A_{0\mathbf{q}} = |F_{\mathbf{q}}|\cos(\varphi_{\mathbf{q}} - \pi q_s)$, we have

$$P(\varphi_{\mathbf{p}} = \pi p_s||F_{\mathbf{p}}|F_{\mathbf{q}}) = 0.5 + 0.5\tanh(|F_{\mathbf{p}}|m_{0\mathbf{p}}/v_{0\mathbf{p}}), \quad (21)$$

where $m_{0\mathbf{p}}$ is given by (16). If $P > 0.5$ then $\varphi_{\mathbf{p}} \simeq \pi p_s$, otherwise $\varphi_{\mathbf{p}} \simeq (\pi p_s + \pi)$.

Relationships (18) and (21) are the desired mathematical formulae for estimating modulus and phase of a structure factor $F_{\mathbf{p}}$ from prior knowledge of $F_{\mathbf{q}}$.

6. The distributions $P[|F_{\mathbf{p}_1}||(F_{\mathbf{p}_j}, j = 2, \ldots, n+1)]$, $P[F_{\mathbf{p}_1}|(F_{\mathbf{p}_j}, j = 2, \ldots, n+1)]$

The relations (18), (19) and (21) answer questions like how $|F_{\mathbf{p}}|$ and $\varphi_{\mathbf{p}}$ may be estimated when $F_{\mathbf{q}}$ is known. If more than one structure factor is *a priori* known, the above question becomes how $|F_{\mathbf{p}_1}|$ and $\varphi_{\mathbf{p}_1}$ may be estimated when $F_{\mathbf{p}_j}$, $j = 2, \ldots, n+1$, are *a priori* known. The problem may be solved if the distributions

$$P[|F_{\mathbf{p}_1}||(F_{\mathbf{p}_j}, j = 2, \ldots, n+1)],$$
$$P[\varphi_{\mathbf{p}_1}|(F_{\mathbf{p}_j}, j = 2, \ldots, n+1)]$$

are derived. The following notation is introduced to make the mathematical formulae shorter:

(*a*) $A_{0\mathbf{p}_j}$, $F_{\mathbf{p}_j}$, $\varphi_{\mathbf{p}_j} \rightarrow A_{0j}$, F_j, φ_j, respectively,

(*b*) the first-order cumulants $K_{10}(\mathbf{p}_j)$ are denoted by K_j,

(*c*) $K_{j_1j_2}$ represents the mixed cumulants between $A_{0\mathbf{p}_{j_1}}$ and $A_{0\mathbf{p}_{j_2}}$,

(*d*) K_{jj} represents the second-order cumulants K_{20} of $A_{0\mathbf{p}_j}$.

We have

$$C(u_1, u_2, \ldots, u_n) = \exp\left(i\sum_{j=1}^{n+1}K_ju_j - 0.5\sum_{j=1}^{n+1}K_{jj}u_j^2 - \sum_{j=2}^{n+1}K_{1j}u_1u_j - \sum_{j_2>j_1=2}^{n+1}K_{j_1j_2}u_{j_1}u_{j_2}\right),$$

where C is the characteristic function of $P(A_{01}, \ldots, A_{0n+1})$ and $u_1, \ldots, u_{n+1}$ are the carrying variables associated with $A_{01}, \ldots, A_{0n+1}$, respectively. Then,

$$P(A_{01}, A_{02}, \ldots, A_{0n+1}) \simeq (2\pi)^{-(n+1)}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}\exp(i\overline{\mathbf{T}}\mathbf{U})\exp(-\tfrac{1}{2}\overline{\mathbf{U}}\mathbf{K}\mathbf{U})\,\mathrm{d}\mathbf{U},$$

where

$$\overline{\mathbf{U}} = (u_1, u_2, \ldots, u_{n+1}),$$

$$\overline{\mathbf{T}} = (d_1, d_2, \ldots, d_{n+1}),$$

$$d_j = K_j - A_{0j},$$

$$\mathbf{K} = \begin{vmatrix} K_{11} & K_{12} & \ldots & K_{1\,n+1} \\ K_{12} & K_{22} & \ldots & K_{2\,n+1} \\ \vdots & \vdots & \ddots & \vdots \\ K_{1\,n+1} & K_{2\,n+1} & \ldots & K_{n+1\,n+1} \end{vmatrix}.$$

$\mathbf{K}$ is a variance–covariance matrix: by definition $(\det\mathbf{K}) \geq 0$. After some calculations, we obtain

$$P(A_{01}, A_{02}, \ldots, A_{0n+1}) \simeq (2\pi)^{-(n+1)/2}(\det\boldsymbol{\lambda})^{1/2}\exp(-\tfrac{1}{2}\overline{\mathbf{T}}\boldsymbol{\lambda}\mathbf{T}), \quad (22)$$

where $\boldsymbol{\lambda} = \mathbf{K}^{-1}$. In a more explicit form, (22) may be written as

$$P(A_{01}, A_{02}, \ldots, A_{0n+1}) \simeq (2\pi)^{-(n+1)/2} (\det \boldsymbol{\lambda})^{1/2} \exp\left(-\tfrac{1}{2} \sum_{j=1}^{n+1} \lambda_{jj} d_j^2 - \sum_{j=2}^{n+1} \lambda_{1j} d_1 d_j - \sum_{j_2 > j_1 = 2}^{n+1} \lambda_{j_1 j_2} d_{j_1} d_{j_2} \right),$$

from which

$$P[A_{01} | (A_{0j}, j = 2, \ldots, n+1)] \simeq [\lambda_{11}/(2\pi)]^{1/2} \exp[-\tfrac{1}{2} \lambda_{11} (A_{01} - M_{01})^2], \quad (23)$$

where

$$M_{01} = K_1 + \lambda_{11}^{-1} \sum_{j=2}^{n+1} \lambda_{1j} d_j \quad (24)$$

is the conditional expected value of A_{01}. Equation (23) is a normal distribution: A_{01} is normally distributed about M_{01} with variance given by

$$V_{01} = \lambda_{11}^{-1}.$$

As in §5, we can rapidly derive from (23) the related distributions and relationships, *i.e.*

(*a*)

$$P[|F_1| | (F_j, j = 2, \ldots, n+1)] = [2/(\pi V_{01})]^{1/2} \exp[-(|F_1|^2 + M_{01}^2)/(2V_{01})] \times \cosh(M_{01}|F_1|/V_{01}), \quad (25)$$

where

$$M_{01} = K_1 + V_{01} \sum_{j=2}^{n+1} \lambda_{1j} d_j, \quad (26)$$

$$d_j = K_j - |F_j| \cos[\varphi_j - \pi p_s(j)].$$

(*b*)

$$|F_1|_{\rm est} = \langle |F_1| | (F_j, j = 2, \ldots, n+1) \rangle = |M_{01}| \Phi[|M_{01}|/(2V_{01})^{1/2}] + (2V_{01}/\pi)^{1/2} \times \exp[-M_{01}^2/(2V_{01})]. \quad (27)$$

If $V_{01} \equiv 0$ then $\Phi[|M_{01}|/(2V_{01})^{1/2}] \equiv \Phi(\infty) = 1$; in such a condition, (27) reduces to

$$\langle |F_1| | (F_j, j = 2, \ldots, n+1) \rangle = |M_{01}|.$$

(*c*)

$$\langle |F_1|^2 | (F_j, j = 2, \ldots, n+1) \rangle = M_{01}^2 + V_{01}. \quad (28)$$

(*d*)

$$P[\varphi_{\mathbf{p}} = \pi p_s | (F_j, j = 2, \ldots, n+1)] = 0.5 + 0.5 \tanh(|F_1| M_{01}/V_{01}). \quad (29)$$

The relations (25)–(29) are the main result of this paper. They can exploit all of the available information, no matter the type of the reflections and their number.

7. The canonical case

The joint probability distribution (22) and the conditional distributions (23), (25) and (29) are valid under quite general conditions. Indeed,

(*a*) the indices of the reflections may be arbitrarily chosen,

(*b*) the value of n is also a free choice; obviously, larger values of n will give rise (in the statistical sense) to more reliable estimates.

However, when n is large, the use of (26) and (29) involves high-order $\mathbf{K}$ and λ matrices. Since the cumulants $K_{11}(j_1, j_2)$ are often non-negligible, the matrices $\mathbf{K}$ are non-diagonal, and their inversion, for very large values of n, may be critical and time consuming. A remarkable simplification of the process may be obtained when (the canonical case) F_{p_1} is a half-integral reflection and F_{p_j}, $j = 2, \ldots, n+1$, are standard reflections (first option); or F_{p_1} is a standard reflection and F_{p_j}, $j = 2, \ldots, n+1$, are half-integral index reflections (second option). In both of the above options, the assumption $K_{j_1 j_2} = 0$ may be made for $(j_1, j_2) \neq (1, j_r)$ and $(j_1, j_2) \neq (j_r, 1)$. Indeed [see the definition of $K_{11}(\mathbf{p}, \mathbf{q})$ in §3]:

(*a*) In the first option, all the three terms in $K_{11}(\mathbf{p}, \mathbf{q})$ which multiply $\Sigma_{11}(\mathbf{p}, \mathbf{q})$ are identically equal to zero when $\mathbf{p}$ and $\mathbf{q}$ are vectors with integral components.

(*b*) In the second option, only two of the terms in $K_{11}(\mathbf{p}, \mathbf{q})$ are equal to zero [indeed $(\mathbf{p} + \mathbf{q})$ and $(\mathbf{p} - \mathbf{q})$ are vectors with integral components]. The third term is different from zero, but the maximum value that its modulus may assume is $[(\sin \pi/2)/(\pi/2)]^6$, attained when $\mathbf{p} = (1/2, 1/2, 1/2)$, $\mathbf{q} = (\pm 1/2, \pm 1/2, \pm 1/2)$. As soon as some of the six components increase, the third term rapidly decreases. Thus, $K_{11}(\mathbf{p}, \mathbf{q})$ are not negligible only when pairs $\mathbf{p}$ and $\mathbf{q}$ are both close to the origin of the reciprocal space.

In the canonical case, therefore, the matrix $\mathbf{K}$ may be reduced to the form

$$\mathbf{K} = \begin{vmatrix} K_{11} & K_{12} & K_{13} & \ldots & K_{1\,n+1} \\ K_{12} & K_{22} & 0 & \ldots & 0 \\ K_{13} & 0 & K_{33} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_{1\,n+1} & 0 & 0 & 0 & K_{n+1\,n+1} \end{vmatrix}$$

where all the elements are vanishing except the diagonal, the first row and the first column entries. In this case, the value of (det $\mathbf{K}$), calculated *via* the Laplace development, is

$$(\det \mathbf{K}) = K_{11}L_{11} + \sum_{j=2}^{n+1} K_{1j}L_{1j},$$

where L_{1j} is the cofactor of K_{1j}. We have

$$L_{11} = \prod_{j=2}^{n+1} K_{jj},$$

$$L_{1j_1} = -K_{1j_1} \prod_{j_2 \neq j_1 = 2}^{n+1} K_{j_2 j_2}.$$

Therefore,

$$(\det \mathbf{K}) = \prod_{j=1}^{n+1} K_{jj} - \sum_{j=2}^{n+1} \left[K_{1j_1}^2 \prod_{j_2 \neq j_1 = 2}^{n+1} K_{j_2 j_2} \right].$$

As an example, if $n = 4$,

$$(\det \mathbf{K}) = (K_{11} \ldots K_{55}) - (K_{12}^2 K_{33} K_{44} K_{55}) - (K_{13}^2 K_{22} K_{44} K_{55}) - \ldots - (K_{15}^2 K_{22} K_{33} K_{44}).$$

Now

$$\lambda_{11}^{-1} = (\det \mathbf{K})/L_{11} = K_{11} - \sum_{j=2}^{n+1} K_{1j}^2 / K_{jj}, \qquad (30)$$

$$\begin{aligned} \lambda_{1j}/\lambda_{11} &= [L_{1j}(\det \mathbf{K})][(\det \mathbf{K})/L_{11}] \\ &= \left[-K_{1j} \prod_{j_2 \neq j_1 = 2}^{n+1} K_{j_2 j_2} \right] \Big/ \prod_{j_2 = 2}^{n+1} K_{j_2 j_2} \\ &= -K_{1j}/K_{jj}. \end{aligned} \qquad (31)$$

The relations (30) and (31) allow a strong simplification in the use of the formulae (25)–(29). Indeed, the inversion of the matrix $\mathbf{K}$ is no longer necessary and the simplified expressions for estimating the parameters necessary to estimate the modulus and of the phase of the reflection $\mathbf{p}$ are

$$M_{01}(\mathbf{p}) = K_1(\mathbf{p}) + \sum_{\mathbf{q}} K_{11}(\mathbf{p}, \mathbf{q}) \times [|F_{\mathbf{q}}| \cos(\varphi_{\mathbf{q}} - \pi q_s) - K_1(\mathbf{q})]/K_2(\mathbf{q}), \qquad (32)$$

$$V_{01} = K_2(\mathbf{p}) - \sum_{\mathbf{q}} [K_{11}^2(\mathbf{p}, \mathbf{q})/K_2(\mathbf{q})]. \qquad (33)$$

8. Applications

Our approach is quite general: any reflection with rational indices (integral indices included) may be estimated *via* any set of reflections, irrespective of the nature of their indices. To reduce the amount of calculations, we restrict our attention to reflections with standard (*i.e.* integral) indices (SI) and to half-integral index (HI) reflections [*e.g.* (1/2, 3/2, 5/2) belongs to the set HI, (3/2, 0, 9/2) or (1/2, 3, 7/2) do not belong to the set].

Table 1. *Estimation of half-integral indices reflections*

Nr is the number of HI reflections which are estimated *via* a number of SI reflections lying in the corresponding NINTO interval.

NINTO	Nr
128–145	157
145–162	226
162–180	241
180–197	250
197–214	265
214–231	263
231–148	252
248–266	260
266–300	1714

The published atomic coordinates of NEWQB (Grigg *et al.*, 1978; $a = 10.3764$, $b = 13.2760$, $c = 14.6548$ Å, $\alpha = 98.712$, $\beta = 95.258$, $\gamma = 93.204°$) were used to calculate SI and HI structure factors up to 1.04 Å resolution. 3675 SI and 3628 HI structure factors (Friedel opposites excluded) were so obtained. The isotropic vibrational parameter $B = 5.0$ was used for all the atoms.

The efficiency of formulae (27) and (29) was checked by estimating HI structure factors from known SI structure factors and *vice versa*. Owing to the probabilistic nature of our approach, we were able to involve in the calculations an arbitrary (the total or a partial) amount of prior information. We decided to estimate each HI structure factor $\mathbf{p}$ from the subset of SI structure factors contained in a given sphere centred on the $\mathbf{p}$ reflection since they are (statistically) the most effective in defining the estimates.. To limit the computing time, the radius of the sphere was assumed to be equal to 0.33 Å^{-1}. The number of SI structure factors constituting the prior information varies with $\mathbf{p}$. The trend may be deduced from Table 1, which may be read as follows: of the 3628 HI reflections, Nr = 1714 (last line of Table 1) were estimated *via* a number of SI structure factors lying between 266 and 300, Nr = 260 were estimated *via* a number of SI structure factors lying between 248 and 266 *etc*. Table 1 suggests that a non-negligible percentage of HI structure factors is estimated *via* a relatively small set of SI structure factors. Such HI reflections are mostly those close to the boundary of the 'measured' Ewald sphere: a reduced number of SI reflections is exploited also when the 0.33 Å^{-1} radius sphere contains Friedel opposites.

The estimates obtained *via* (27) are ranked in Table 2 according to $|F_1|_{\rm est}/V_{01}^{1/2}$: Nr is the number of estimates with $|F_1|_{\rm est}/V_{01}^{1/2} >$ ARG, R is the corresponding residual value given by

$$R = \left(\sum_{\mathbf{p}} \big| |F_{\mathbf{p}}| - |F_{\mathbf{p}}|_{\rm est} \big| \right) \Big/ \left(\sum_{\mathbf{p}} |F_{\mathbf{p}}| \right),$$

where $F_{\mathbf{p}}$ is calculated from the published atomic coordinates, $|F_{\mathbf{p}}|_{\rm est}$ is the corresponding estimate provided by (27). Table 2 shows that, in spite of the limited amount of prior information, accurate estimates of the moduli $|F_{\mathbf{p}}|$

Table 2. *Estimation of half-integral indices reflections*

Nr is the number of estimates with $|F_1|_{est}/V_{01}^{1/2} >$ ARG, R is the corresponding residual value given by $R = (\Sigma_{\mathbf{p}}||F_{\mathbf{p}}| - |F_{\mathbf{p}}|_{est}|)/(\Sigma_{\mathbf{p}}|F_{\mathbf{p}}|)$.

ARG	Nr	R
0.80	3628	0.37
0.97	2594	0.33
1.23	2005	0.31
1.50	1607	0.28
1.77	1288	0.26
2.03	1058	0.25
2.30	814	0.23
2.57	652	0.22
2.83	534	0.21
3.10	421	0.19
3.37	322	0.18
3.63	253	0.17

Table 3. *Estimation of half-integral reflections*

Nr is the number of estimates with $|F_{\mathbf{p}}|M_{0\mathbf{p}}/V_{01}^{1/2} >$ ARG and % is the percentage of the correct phase estimates.

ARG	Nr	%
0.0	3628	85
0.5	3486	86
1.0	3349	87
2.5	3221	89
4.0	3100	90
5.5	2991	91
7.0	2873	93
9.5	2472	96
12.0	2180	96
14.5	1996	97
17.0	1824	97
30.0	1071	99
50.0	688	100
60.0	573	100
70.0	469	100

Table 4. *Estimation of integral indices reflections*

Nr is the number of SI reflections which are estimated *via* a number of HI reflections lying in the corresponding NINTO interval.

NINTO	Nr
118–136	127
136–154	342
154–173	232
173–191	256
191–209	252
209–227	265
227–245	233
245–263	258
263–300	1709

are always obtained, except for small values of $|F_1|_{est}/V_{01}^{1/2}$. The estimates improve when the radius of the sphere around $\mathbf{p}$ is enlarged (the results are not shown for brevity).

The efficiency of (29) for the sign determination may be deduced from Table 3, where Nr is the number of estimates with $|F_{\mathbf{p}}|M_{0\mathbf{p}}/V_{01}^{1/2} >$ ARG and % is the percentage of the correct phase estimates. Tables 2 and 3 suggest that wrong estimates (of both modulus and phase) mostly occur for small moduli $|F_{\mathbf{p}}|$.

Table 5. *Estimation of integral indices reflections*

Nr is the number of estimates with $|F_1|_{est}/V_{01}^{1/2} >$ ARG, R is the corresponding residual value given by $R = (\Sigma_{\mathbf{p}}||F_{\mathbf{p}}| - |F_{\mathbf{p}}|_{est}|)/(\Sigma_{\mathbf{p}}|F_{\mathbf{p}}|)$.

ARG	Nr	R
0.80	3675	0.46
0.97	2578	0.43
1.23	1957	0.40
1.50	1577	0.38
1.77	1279	0.36
2.03	1014	0.34
2.30	823	0.33
2.57	666	0.32
2.83	523	0.32
3.10	410	0.31
3.37	217	0.31
3.63	250	0.32

Table 6. *Estimation of integral reflections*

Nr is the number of estimates with $|F_{\mathrm{p}}|M_{0\mathrm{p}}/V_{01}^{1/2} >$ ARG and % is the percentage of the correct phase estimates.

ARG	Nr	%
0.0	3675	79
0.5	3537	80
1.0	3407	81
2.5	3264	82
4.0	3154	83
5.5	3031	84
7.0	2931	84
9.5	2501	86
12.0	2196	88
14.5	1952	90
17.0	1766	91
30.0	1072	93
50.0	682	93
60.0	558	93
70.0	467	93

Analogous results have been obtained when SI reflections were estimated *via* HI reflections (as calculated from published atomic coordinates). They are summarized in Tables 4, 5 and 6. It may be observed that the results in Tables 5 and 6 are slightly poorer than those quoted in Tables 2 and 3. This is because, in the second option of the canonical case, the matrix **K** is not identical to but only approximated by the simpler form described in §7.

9. Conclusions

The method of joint probability distribution functions has been generalized to involve structure factors with rational indices. Formulae estimating modulus and phase of a given structure factor (with rational or integral indices) have been derived, provided the moduli and/or the phases of other structure factors are *a priori* known. The restraint that the indices of the reflections have to constitute structure invariants or seminvariants is no longer necessary.

To understand the possible applications of this paper, a final remark might be necessary. Rational indices have found various applications in crystallography: they are involved in the rotation function for molecular replacement methods, in the solution of problems involving non-crystallographic symmetry (for brevity we do not quote the wide literature on these two subjects), in the determination of the signs of the centrosymmetric structure factors (Boyes-Watson *et al.*, 1947). In recent years, a revival in the interest of reflections with non-integral indices has occurred, originated by a seminal paper by Ramachandran (1969), who investigated the possible use of the Hilbert transform in crystallography. His equations involve unknown derivatives: a solution of the problem was proposed by Mishnev (1993) through the application of the Shannon (1949) sampling theorem, and applications were made by Zanotti *et al.* (1996).

The present paper is the first to describe a probabilistic approach to the use of the reflections with rational indices. Its connections with the algebraic approaches described in the above-mentioned literature will not be discussed here, and is postponed to the next paper of this series, where the case $P1$ will be illustrated.

APPENDIX *A*

Let us suppose that in the space group $P\bar{1}$ the primitive variables, $x_j, y_j, z_j, j = 1, \ldots, N$, are independently and uniformly distributed in the interval (0, 1). Then

$$\begin{aligned} m_{11} &= \langle A_{0\mathbf{p}} A_{0\mathbf{q}} \rangle \\ &= 4 \sum_{j=1}^{N/2} f_j(\mathbf{p}) f_j(\mathbf{q}) \langle \cos(\pi p_s - 2\pi \mathbf{p} \cdot \mathbf{r}_j) \\ &\quad \times \cos(\pi q_s - 2\pi \mathbf{q} \cdot \mathbf{r}_j) \rangle + 4 \sum_{j_1 \neq j_2 = 1}^{N/2} f_j(\mathbf{p}) f_j(\mathbf{q}) \\ &\quad \times \langle \cos(\pi p_s - 2\pi \mathbf{p} \cdot \mathbf{r}_j) \cos(\pi q_s - 2\pi \mathbf{q} \cdot \mathbf{r}_j) \rangle \\ &= \left[4 \sum_{j=1}^{N/2} f_j(\mathbf{p}) f_j(\mathbf{q}) \right] \{ 0.5 c_{\mathbf{p}+\mathbf{q}} \cos[\pi(p_s + q_s)] \\ &\quad + 0.5 s_{\mathbf{p}+\mathbf{q}} \sin[\pi(p_s + q_s)] + 0.5 c_{\mathbf{p}-\mathbf{q}} \cos[\pi(p_s - q_s)] \\ &\quad + 0.5 s_{\mathbf{p}-\mathbf{q}} \sin[\pi(p_s - q_s)] \} \\ &\quad + \left[4 \sum_{j_1 \neq j_2 = 1}^{N/2} f_{j_1}(\mathbf{p}) f_{j_2}(\mathbf{q}) \right] [c_{\mathbf{p}} \cos(\pi p_s) + s_{\mathbf{p}} \sin(\pi p_s)] \\ &\quad \times [c_{\mathbf{q}} \cos(\pi q_s) + s_{\mathbf{q}} \sin(\pi q_s)], \end{aligned} \quad (34)$$

where (see §I.12)

$$\begin{aligned} c_{\mathbf{p}} &= \langle \cos(2\pi \mathbf{p} \cdot \mathbf{r}_j) \rangle \\ &= c_{p_1} c_{p_2} c_{p_3} - c_{p_1} s_{p_2} s_{p_3} - s_{p_1} s_{p_2} c_{p_3} - s_{p_1} c_{p_2} s_{p_3}, \end{aligned}$$

$$\begin{aligned} s_{\mathbf{p}} &= \langle \sin(2\pi \mathbf{p}.\mathbf{r}_j) \rangle \\ &= s_{p_1} c_{p_2} c_{p_3} - s_{p_1} s_{p_2} s_{p_3} - c_{p_1} s_{p_2} c_{p_3} - c_{p_1} c_{p_2} s_{p_3}. \end{aligned}$$

Since

(*a*)

$$\begin{aligned} \left[\sum_{j=1}^{N/2} f_j(\mathbf{p}) \right] \left[\sum_{j=1}^{N/2} f_j(\mathbf{q}) \right] &= [0.5 \Sigma_1(\mathbf{p})][0.5 \Sigma_1(\mathbf{q})] \\ &= \sum_{j=1}^{N/2} f_j(\mathbf{p}) f_j(\mathbf{q}) + \sum_{j_1 \neq j_2 = 1}^{N/2} f_{j_1}(\mathbf{p}) f_{j_2}(\mathbf{q}) \\ &= 0.5 \Sigma_{11}(\mathbf{p}, \mathbf{q}) + \sum_{j_1 \neq j_2 = 1}^{N/2} f_{j_1}(\mathbf{p}) f_{j_2}(\mathbf{q}), \end{aligned}$$

(*b*)

$$c_{2\mathbf{p}} \cos(2\pi p_s) + s_{2\mathbf{p}} \sin(2\pi p_s) = c_{p_1} c_{p_2} c_{p_3},$$

we can rearrange (34) into the following expression:

$$\begin{aligned} m_{11} &= \Sigma_{11}(\mathbf{p}, \mathbf{q}) \{ c_{(p_1+q_1)/2} c_{(p_2+q_2)/2} c_{(p_3+q_3)/2} \\ &\quad + c_{(p_1-q_1)/2} c_{(p_2-q_2)/2} c_{(p_3-q_3)/2} \} \\ &\quad + [\Sigma_1(\mathbf{p}) \Sigma_1(\mathbf{q}) - 2 \Sigma_{11}(\mathbf{p}, \mathbf{q})] \\ &\quad \times (c_{p_1/2} c_{p_2/2} c_{p_3/2} c_{q_1/2} c_{q_2/2} c_{q_3/2}) \\ &= \Sigma_{11}(\mathbf{p}, \mathbf{q}) [c_{(p_1+q_1)/2} c_{(p_2+q_2)/2} c_{(p_3+q_3)/2} \\ &\quad + c_{(p_1-q_1)/2} c_{(p_2-q_2)/2} c_{(p_3-q_3)/2} \\ &\quad - 2 (c_{p_1/2} c_{p_2/2} c_{p_3/2} c_{q_1/2} c_{q_2/2} c_{q_3/2}) \\ &\quad + \Sigma_1(\mathbf{p}) \Sigma_1(\mathbf{q}) (c_{p_1/2} c_{p_2/2} c_{p_3/2} c_{q_1/2} c_{q_2/2} c_{q_3/2})], \end{aligned}$$

which is the required expression for m_{11}. The value of K_{11} quoted in §3 easily follows from the relation $K_{11} = m_{11} - m_{10} m_{01}$.

APPENDIX *B*

We collect here, for the convenience of the reader, the main integration formulae used in this paper (Gradshteyn & Ryzhik, 1965),

$$\int_{-\infty}^{\infty} \exp(-px^2 + 2qx)\, \mathrm{d}x = (\pi/p)^{1/2} \exp(q^2/p), \quad (35)$$

$$\int_{-\infty}^{\infty} x \exp(-px^2 + 2qx)\, \mathrm{d}x = (q/p)(\pi/p)^{1/2} \exp(q^2/p), \quad (36)$$

$$\int_{0}^{\infty} \exp(-px^2 + 2qx)\, \mathrm{d}x = 2^{-1}(\pi/p)^{1/2} \exp(q^2/p) \times [1 + \Phi(q/p^{1/2})], \quad (37)$$

where

$$\Phi(x) = 2\pi^{-1/2} \int_0^x \exp(-t^2)\, \mathrm{d}t$$

is the probability integral;

$$\int_{-\infty}^{\infty} x \exp(-px^2 + 2qx)\, \mathrm{d}x = (2p)^{-1} + [q/(2p)](\pi/p)^{1/2} \exp(q^2/p) \times [1 + \Phi(q/p^{1/2})], \quad (38)$$

$$\int_0^{\infty} x \exp(-px^2) \cosh(qx)\, \mathrm{d}x = [q/(4p)](\pi/p)^{1/2} \exp(q^2/4p) \Phi[q/(2p^{1/2})] + (2p)^{-1}, \quad (39)$$

$$\int_0^{\infty} x \exp(-px^2) \cosh(qx)\, \mathrm{d}x = \pi^{1/2}(2p + q^2) \exp(q^2/4p)/(8p^2 p^{1/2}). \quad (40)$$

References

Boyes-Watson, J., Davidson, E. & Perutz, M. F. (1947). *Proc. R. Soc. London Ser. A*, **191**, 83–132.

Giacovazzo, C. & Siliqi, D. (1998). *Acta Cryst.* A**54**, 957–970.

Giacovazzo, C., Siliqi, D., Carrozzini, B., Guagliardi, A. & Moliterni, A. G. (1999). *Acta Cryst.* A**55**, 314–321.

Gradshteyn, I. S. & Ryzhik, I. M. (1965). *Table of Integrals, Series and Products.* New York: Academic Press.

Grigg, R., Kemp, J., Sheldrick, G. M. & Trotter, J. (1978). *J. Chem. Soc. Chem. Commun.* pp. 109–111.

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

Mishnev, A. F. (1993). *Acta Cryst.* A**49**, 159–161.

Ramachadran, G. N. (1969). *Mater. Res. Bull.* **4**, 525–534.

Shannon, C. E. (1949). *Proc. Inst. Radio Eng. NY*, **37**, 10–21.

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

Zanotti, G., Fogale, F. & Capitani, G. (1996). *Acta Cryst.* A**52**, 757–765.