

with that used in lattice dynamics for defining Brillouin zones.

#### References

- JANNER, A. & JANSSEN, T. (1979). *Physica (Utrecht)*, **99A**, 47-76.
- JANNER, A., JANSSEN, T. & DE WOLFF, P. M. (1983). *Acta Cryst.* **A39**, 658-666.
- WOLFF, P. M. DE (1974). *Acta Cryst.* **A30**, 777-785.
- WOLFF, P. M. DE, JANSSEN, T. & JANNER, A. (1981). *Acta Cryst.* **A37**, 625-636.
- YAMAMOTO, A., JANSSEN, T., JANNER, A. & DE WOLFF, P. M. (1985). *Acta Cryst.* **A41**, 528-530.

*Acta Cryst.* (1989). **A45**, 63-75

## Statistical Descriptors in Crystallography

### Report of the International Union of Crystallography Subcommittee on Statistical Descriptors\*

BY D. SCHWARZENBACH (Chairman), *Institut de Cristallographie, University of Lausanne, BSP Dorigny, CH-1015 Lausanne, Switzerland*, S. C. ABRAHAMS (*ex officio*, IUCr Commission on Crystallographic Nomenclature), *AT&T Bell Laboratories, 600 Mountain Ave, Murray Hill, NJ 07974, USA*, H. D. FLACK, *Laboratoire de Cristallographie aux Rayons X, University of Geneva, 24 Quai Ernest Ansermet, CH-1211 Genève 4, Switzerland*, W. GONSCHOREK, *Institut für Physikalische Chemie der TH Darmstadt, Aussenstelle in KFA/ ZFR, Postfach 913, D-5170 Jülich, Federal Republic of Germany*, TH. HAHN (*ex officio*, IUCr Commission on Crystallographic Nomenclature), *Institut für Kristallographie, RWTH, Templergraben 55, D-5100 Aachen, Federal Republic of Germany*, K. HUML (*ex officio*, IUCr Commission on Crystallographic Computing), *Laboratory of X-ray Polymer Structure Analysis, Institute of Macromolecular Chemistry, Czech Academy of Science, 162 06 Praha 6, Czechoslovakia*, R. E. MARSH, *Chemistry Department, California Institute of Technology, Pasadena, CA 91125, USA*, E. PRINCE, *National Bureau of Standards, Reactor Radiation Division, Gaithersburg, MD 20899, USA*, B. E. ROBERTSON, *Department of Physics and Astronomy, University of Regina, Regina, Saskatchewan S4S 0A2, Canada*, J. S. ROLLETT, *Oxford University Computing Laboratory, 8-11 Keble Road, Oxford OX1 3QD, England* AND A. J. C. WILSON (*ex officio*, IUCr Commission on Crystallographic Nomenclature), *University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England*

(Received 28 March 1988; accepted 2 September 1988)

#### Abstract

The Subcommittee has attempted to elucidate the nature of problems encountered in the definition and use of statistical descriptors as applied to crystallography and to propose procedural improvements. The report contains (a) a dictionary of statistical terms established for use by experimentalists; (b) a description of the statistical basis for refinement procedures; (c) sections dealing with defects in the physical model used for refinement, and with the choice and significance of weighting schemes; and (d) recommendations, some of which may be readily implemented, whilst others may require a long-term effort to bring them into general use.

#### Introduction

A result of several discussions at the XIII International Congress of Crystallography in Hamburg,

1984, particularly those arising at a microsymposium devoted to crystallographic statistics, was a renewed recognition of the wide nonuniformity in use and nomenclature of many statistical methods applied to crystallography. The Commission on Crystallographic Nomenclature addressed this problem soon after the Congress had ended and agreed that an attempt should be made to improve the situation. Accordingly, a Subcommittee on Statistical Descriptors in Crystallography was appointed in early 1985 with its terms of reference 'to examine the validity of current statistical approaches used in estimating the variances in crystallographic quantities and to make recommendations for an improved methodology that rests securely on sound modern statistical theory and that can be widely adopted by the crystallographic community'.

Vigorous correspondence within the Subcommittee resulted in a series of draft reports that gradually evolved toward general consensus. An intermediate report was presented orally at an Open Meeting of the Commission during the XIV International Congress of Crystallography in Perth 1987.

Problems arising from the interface between the mathematical theory of probability and statistics and

\* Appointed 27 February 1985 as a Subcommittee of the IUCr Commission on Crystallographic Nomenclature. Following a review by the Chairmen of all relevant IUCr Commissions, the Final Report was accepted on 9 May 1988 by the Commission and on 2 September 1988 by the Executive Committee.

the practising experimentalist are not confined to procedural issues. One cannot tackle such problems without realizing their important philosophical component originating from the need to justify the lack of rigour and the unavoidable approximations in the treatment of experimental data. For this reason, the attention of the non-specialist is drawn to the two main interpretations of probability used by statisticians. In the **frequentist** point of view, the probability of an event is taken to be equal to the limit of the relative frequency of the chosen event with respect to all possible events as the number of trials goes to infinity. The appeal of the frequentist approach for physical scientists lies in the apparent objectivity of its treatment of data. Almost all textbooks of statistics written for physical scientists follow this approach (e.g. Hamilton, 1964). On the other hand, the **Bayesian** approach extends the interpretation of probability to include degrees of belief or knowledge in propositions. We pass from the probability of events (frequentist) to the probability of propositions (Bayesian). Nevertheless the axioms used to define the mathematical properties of probability remain unchanged. Consequently many of the statistical procedures of the two approaches are identical, apart from some changes of emphasis. The Bayesian approach is very scantily mentioned in textbooks for physical scientists (see, however, French, 1978; Box & Tiao, 1973). The frequentist school reproaches the Bayesians for their apparent lack of objectivity. The Bayesians consider that objectivity in statistics is illusory, noting that everything is interpreted through the use of preconceived models. Most physical scientists argue in terms of frequentist concepts but implicitly use a more subjective or Bayesian touch in dealing with their experimental data. Often-encountered expressions like 'the probable value of a parameter' are meaningful only in the Bayesian framework (French & Oatley, 1982).

### Definition of statistical terms

**Accuracy:** The closeness of agreement between the value of an *estimate*, derived from a physical measurement, and the true value of the quantity (measurand) estimated. The reference to the true value implies that in practice accuracy cannot be exactly evaluated. The terms *accuracy* and *precision* must not be confused. Experimental science endeavours to gain insight into **physical reality** (or truth) through interpretation of measurements using *models*. This is based on the implicit assumption that a bad agreement between observations and the corresponding calculated model quantities indicates inaccuracy (see *Goodness of fit*); the converse of this proposition, that good agreement between observations and model quantities indicates accuracy, is thereby not implied. A model that does not take into account all available evidence and prior

experience may give apparently precise, but inaccurate (wrong) results.

**Average:** The average of a set of values  $\{x_i\}$ ,  $1 \leq i \leq n$ , is defined by  $\bar{x} = (\sum_i x_i)/n$ . If the  $\{x_i\}$  are a *sample* of  $n$  independent observations of a single quantity  $x$  distributed according to a *probability density function*  $p(x)$  with *mean*  $\mu$  and *variance*  $\sigma^2$ , then  $\bar{x}$  is a minimum-variance unbiased *estimate* of  $\mu$ ;

$$s^2 = \left[ \sum_i (x_i - \bar{x})^2 \right] / (n - 1)$$

is an unbiased estimate of  $\sigma^2$ . The variance of the probability density function of  $\bar{x}$  is  $\sigma^2/n$  and an unbiased estimate is obtained from  $s^2/n$ . These estimates do not require a complete knowledge of  $p(x)$  to be available. Using a set of *weights*  $\{w_i\}$ , the **weighted average** is defined by  $\bar{x}_w = \sum_i w_i x_i / \sum_i w_i$ . If the  $\{w_i\}$  do not depend on the  $\{x_i\}$ , then  $\bar{x}_w$  is an unbiased estimate of  $\mu$ . The weighted average finds its use in cases where the  $n$  observations are drawn from populations of identical mean but differing variances.

**Bayesian:** An interpretation of probability developed from **Bayes's theorem** [Bayes, 1763; see *Probability density function*, equation (17) and *Introduction*]. Bayes's theorem itself is accepted by either frequentist or Bayesian statisticians. However, the assignment of prior probabilities in Bayesian inference has caused much discussion. **Bayes's postulate** assumes an equipartition of ignorance and states that in the absence of information to the contrary all prior probabilities are assumed to be equal. The work of Jaynes (1983), on the other hand, shows that by using the concepts of group invariance uninformative prior probabilities may be obtained which are far from the uniform distributions suggested by a casual appreciation of Bayes's postulate. At the present time, many Bayesian statisticians would contend that the search for uninformative prior probabilities is misguided since some prior knowledge of the system studied is always available.

**Bias:** An *estimator* of a statistical quantity is biased if the *expected value* of the quantity is not equal to the true value. In the physical sciences, bias is usually considered as synonymous with *systematic error*. In statistics, it is sometimes restricted to a particular type of systematic error arising from the mathematical model applied to the observations. Any non-linear operation or *model* can result in a bias, as, for example, taking the square root of an intensity when computing structure amplitudes. The bias is then due to the fact that the *expected value* of a function  $f(x)$  is in general not a simple function of the expected value of  $x$ :

$$E[f(x)] \neq f[E(x)]. \quad (1)$$

The equality will hold for any distribution of  $x$  if  $f(x)$  is a linear function of  $x$ . Biases in this restricted sense are proportional to the *variances* due to *random errors* of the observations (Wilson, 1976b).

*Conditional probability density function*: see *Probability density function*.

*Correlation*: see *Moment*.

*Covariance*: see *Moment*.

*Cumulative distribution function (c.d.f.)*: The function  $P(x)$  obtained by integration of the continuous probability density function (p.d.f.)  $p(t)$ :

$$P(x) = \int_{-\infty}^x p(t) dt. \quad (2)$$

*Degrees of freedom*: When  $m$  parameters are estimated from  $n$  observations ( $n \geq m$ ), the quantity  $n - m$  is called the number of degrees of freedom for error.

*Deviance*: If  $O_j$  are the observed values, and  $C_j$  the corresponding calculated model values of  $n$  quantities, the differences  $d_j = O_j - C_j$  are called **deviates**. The definition holds for any given choice of model parameters (see *residual*). The deviance between the calculated and observed quantities is

$$D_o = \sum_{j=1}^n (O_j - C_j)^2 = \sum_{j=1}^n d_j^2. \quad (3)$$

The **weighted deviance** is

$$D_w = \sum_{j=1}^n \sum_{k=1}^n w_{jk} (O_j - C_j)(O_k - C_k) = \mathbf{d}^T \mathbf{W} \mathbf{d}, \quad (4)$$

where  $\mathbf{d}$  is an  $n$  vector of deviates. The positive-definite  $n \times n$  weight matrix  $\mathbf{W}$  may be written as a product  $\mathbf{W} = \mathbf{B}^T \mathbf{B}$ . The matrix  $\mathbf{B}$  is not in general uniquely defined and may be specified, for example, as an upper-triangular matrix, or alternatively as a symmetric matrix. Equation (4) then becomes  $D_w = \mathbf{d}^T \mathbf{B}^T \mathbf{B} \mathbf{d}$ , where  $\mathbf{B} \mathbf{d}$  is the vector of **weighted deviates**. If correlation terms of  $\mathbf{W}$  are assumed to be negligible,  $\mathbf{W}$  becomes a diagonal matrix, and the quantity commonly refined in least squares is obtained:

$$D_w = \sum_{j=1}^n w_j (O_j - C_j)^2. \quad (4')$$

The weighted deviates are then  $d_j = w_j^{1/2} (O_j - C_j)$ . If the weights are the reciprocals of the *variances* of the observed quantities,  $w_j = \sigma_j^{-2}$ , or more generally  $\mathbf{W} = \mathbf{V}^{-1}$ , where  $\mathbf{V}$  is the *variance-covariance matrix* of the observations, the weighted deviance is often called the **scaled deviance**. For *maximum likelihood*, the scaled deviance is  $-2 \ln L_{\max}$ .

*Deviate*: see *Deviance*.

*Durbin-Watson  $d$  statistic*:  $d$  quantifies the serial correlation of least-squares deviates. In its original form it is defined by

$$d = \sum_{j=2}^N (d_j - d_{j-1})^2 / \sum_{j=1}^N d_j^2, \quad (5)$$

where  $d_j = O_j - C_j$  is the *deviate*.  $d$  takes values  $0 <$

$d < 4$ . For no serial correlation a value close to 2 is expected. With positive serial correlation, adjacent deviates tend to have the same sign and  $d$  becomes less than 2, whereas with negative serial correlation (alternating signs of deviates)  $d$  takes values larger than 2. Tables of values for testing  $d$  are given by Durbin & Watson (1950, 1951, 1971) and a convenient approximation to the tables based on the normal distribution is provided by Theil & Nagar (1961). Clearly, the sequencing of the observations is important in the evaluation of  $d$ . For a least-squares fit where the value(s) of  $d$  are significantly different from 2, the estimates of the variances and covariances of the parameters can be grossly in error, being either too large or too small. The serial correlation of the deviates may arise from a time dependence of the observations [e.g. time series; Flack, Vincent & Vincent (1980)], from an experimental correlation of the observations [e.g. gas electron diffraction; Morino, Kuchitsu & Murata (1965); Murata & Morino (1966)] or, most importantly in crystallography, from inadequacy in the physical model used in the least-squares fit [e.g. Rietveld analysis; Hill & Flack (1987)]. The most general form of  $d$  is given by

$$d = \frac{\mathbf{d}^T \mathbf{B}^T \mathbf{P} \mathbf{B} \mathbf{d}}{\mathbf{d}^T \mathbf{W} \mathbf{d}}, \quad (5')$$

where  $\mathbf{d}$ ,  $\mathbf{W}$  and  $\mathbf{B}$  are defined in the section *Deviance*,  $\mathbf{P}$  is an  $n \times n$  matrix with  $P_{jk} = 2$  for  $|j - k| = 0$ ,  $P_{jk} = -1$  for  $|j - k| = 1$ , and  $P_{jk} = 0$  for  $|j - k| > 1$ , and  $\mathbf{B}^T \mathbf{B} = \mathbf{W}$ .

*Estimate*: A value of a physical quantity obtained from the observations by use of an *estimator*. The quantities usually estimated in crystallography are lattice parameters and atomic positional and displacement parameters.

*Estimated standard deviation (e.s.d.)*: An estimate of the *standard deviation*, or square root of the *variance*, of a *probability density function*. Methods used for obtaining e.s.d.'s of diffraction intensities may take into account quantum counting statistics, the variations of periodically measured check reflections, and the scatter among symmetry-equivalent reflections. Such methods are part of the *model*, and not of the observations.

*Estimator*: A mathematical expression (function) leading from the observations to an *estimate* of the value of a physical quantity. An estimator is **unbiased** if its *expected value* is equal to the true value of the quantity (see *Model*). In the presence of *bias* in the restricted sense which is proportional to the variances due to *random errors* of the observations, the expected value of the estimator changes with observation time, and converges to the unbiased value (in the restricted sense) for an eternal immutable experiment.

*Expected value*: Defined as

$$E[f(x)] = \int_{-\infty}^{+\infty} f(x) p(x) dx \quad (6)$$

for the function,  $f(x)$ , of a *random variable*  $x$  whose *probability density function* is  $p(x)$ .  $E[f(x)]$  is not a function of  $x$ , but it does depend on  $p(x)$ . **Expectation** is a synonym of expected value. The expected value of  $x$ ,  $E(x)$ , is called the **mean** of the probability density function and is often denoted by  $\mu$ .

**Gaussian**: see *Normal probability density function*.

**Goodness of fit**: A measure of the extent to which calculated model values  $C_j$  of a set of  $n$  quantities (e.g. X-ray intensities) approach the observed values  $O_j$ . In the statistical literature, the term denotes a class of hypothesis tests. For a crystallographic least-squares refinement with weight matrix  $W$ , it is defined as the square root of the weighted *deviance* divided by its *expected value*:

$$S^2 = \mathbf{d}^T \mathbf{W} \mathbf{d} / E(\mathbf{d}^T \mathbf{W} \mathbf{d}). \quad (7)$$

The deviation of  $S^2$  from unity is a measure of the validity of the *model* used to compute  $C_j$ , and of the estimate of the *variance-covariance matrix*  $V$  of the observations used to calculate  $E(\mathbf{d}^T \mathbf{W} \mathbf{d})$ . If, and only if, the weight matrix  $W$  in the refinement is chosen to be the inverse of  $V$ ,  $W = V^{-1}$ , then  $E(\mathbf{d}^T \mathbf{W} \mathbf{d}) = n - m$ , regardless of the form of the p.d.f. (Wilson, 1980b);  $m$  is the number of variables in the model, and  $n - m$  the number of *degrees of freedom*. For a diagonal matrix  $V$ ,  $S^2$  then becomes

$$S^2 = (n - m)^{-1} \sum_{j=1}^n \sigma_j^{-2} (O_j - C_j)^2. \quad (7')$$

Further, for the  $W = V^{-1}$  weighting scheme, if the *deviates* are *normally distributed* (i.e. with *Gaussian* joint p.d.f.), the value of  $S^2$  that will be exceeded in  $100\alpha\%$  of replications is given by

$$(S^2)_\alpha = \chi_{n-m, \alpha}^2 / (n - m), \quad (8)$$

where  $\chi_{n-m, \alpha}^2$  is the  $100\alpha\%$  point of the  $\chi^2$  distribution. See Abrahams (1969) for a fuller discussion, including calculation of the expected range of  $S^2$  at a given confidence level. It is common practice among crystallographers refining parameters by least squares to multiply the corresponding *estimated standard deviations* by  $S$ . This questionable practice is equivalent to the assumption that a lack of fit is due entirely to an underestimate of the variances of the observations, whose relative values have been correctly assigned.

**Maximum likelihood**: If the  $m$  vector  $\mathbf{x}$  is a set of *parameters*, and the  $n$  vector  $\mathbf{y}$  is a set of observations, then the *conditional probability density function*  $p_C(\mathbf{y}|\mathbf{x})$  can be considered to be a density function for  $\mathbf{x}$  giving the likelihood of observing  $\mathbf{y}$ . It is written  $L(\mathbf{x}|\mathbf{y})$ . The method of maximum likelihood finds the maximum of  $L$  as a function of  $\mathbf{x}$ . It is often simpler to work with the natural logarithm,  $\ln L$ . Maximum likelihood is equivalent to least squares for a *normal* distribution of errors and to Bayesian estimation using a uniform prior distribution.

**Mean**: term used for the *expected value* of  $x$ ,  $E(x)$ , of a *probability density function*  $p(x)$ .

**Model**: Conjecture about physical reality used to interpret the observations. An *estimator* is constructed using the mathematical formulation of the model. In crystallography, the observations are usually integrated intensities and associated backgrounds, but may also include other information, e.g. crystal dimensions. The standard model is kinematical X-ray or neutron diffraction by a crystal composed of spherical atoms or point nuclei undergoing harmonic displacements. The term 'corrected observations' leads to confusion. Any correction applied to the observations is part of the model, and results in a quantity which has not been observed. This includes all data-reduction procedures to obtain structure amplitudes, including absorption and Lorentz-polarization corrections and averaging of symmetry-equivalent data. Thus, structure amplitudes are not observed quantities. For the sake of computational efficiency, corrections without adjustable parameters may be applied to the observations in a data-reduction procedure. However, any non-linear process, e.g. taking the square root, may result in additional *bias*.

**Moment**: The *expected value* of the  $n$ th power,  $x^n$ , of the *random variable*  $x$  is called the  $n$ th moment of the *probability density function* (p.d.f.)  $p(x)$ :

$$E(x^n) = \int_{-\infty}^{+\infty} x^n p(x) dx. \quad (9)$$

The first moment or **mean** is commonly denoted by  $E(x) = \mu$ . The second moment about the mean,  $E[(x - \mu)^2]$ , is the **variance** of  $p(x)$  and is commonly denoted by  $\sigma^2$ . In a space of  $n$  dimensions, the mean of the joint *probability density function*  $p_J(\mathbf{x})$  is an  $n$  vector  $\mathbf{E}(\mathbf{x})$  with elements  $\mu_i = E(x_i)$  equal to the mean of the marginal p.d.f. of  $x_i$ . The  $n \times n$  **variance-covariance matrix** (tensor of rank 2 in  $n$  dimensions) is defined by the second moments about the mean:

$$\begin{aligned} V_{ij} &= E[(x_i - \mu_i)(x_j - \mu_j)] \\ &= \int (x_i - \mu_i)(x_j - \mu_j) p_J(\mathbf{x}) d^n(\mathbf{x}). \end{aligned} \quad (10)$$

For  $i = j$ , this is equal to the variance of the marginal p.d.f. of  $x_i$ ,  $V_{ii} = \sigma_i^2$ . The **correlation matrix** is defined by

$$\rho_{ij} = V_{ij} / (\sigma_i \sigma_j), \quad (11)$$

$-1 \leq \rho_{ij} \leq 1$  and  $\rho_{ii} = 1$ . The set of  $M$ th moments of  $p_J(\mathbf{x})$ ,  $E[x_1^{m(1)} x_2^{m(2)} \dots x_n^{m(n)}]$  with  $m(1) + m(2) + \dots + m(n) = M$ , transforms as a totally symmetric tensor of rank  $M$  in  $n$  dimensions.

**Normal (Gaussian) probability density function**: A *probability density function* of the quantity  $x$  with standard deviation  $\sigma$  about the mean  $\mu$ , given by

$$p(x) = \sigma^{-1} (2\pi)^{-1/2} \exp\{-\frac{1}{2}[(x - \mu)/\sigma]^2\}. \quad (12)$$

The normal probability density function can originate

from addition of a large number of small independent errors, each with its own non-normal distribution, but occurrence of a normal p.d.f. does not imply this underlying structure. The **standard normal deviate**  $z = (x - \mu)/\sigma$  has a normal distribution with zero mean and unit standard deviation, when  $x$  is distributed according to (12).

**Normal probability plot:** A graphical procedure in which the differences between two independent sets of measurements, or those between experiment and theory, are analyzed in terms of a *normal (Gaussian) p.d.f.* The ordered experimental **normal deviates** or *order statistics* such as

$$\delta m_i = [O(1)_i - kO(2)_i] / [\sigma^2(1)_i + k^2\sigma^2(2)_i]^{1/2},$$

where  $O(1)$  and  $O(2)$  are independent observations with variances  $\sigma^2(1)$  and  $\sigma^2(2)$  of the same quantity obtained in separate experiments and  $k$  is a scale factor, or  $\delta R_i = (O_i - C_i)/\sigma_i$  (see *Goodness of fit*) are plotted against the ordered standard normal deviates. A resulting normal probability plot that is linear, with zero intercept and unit slope, shows that the experimental deviates are normally distributed (Abrahams & Keve, 1971; Hamilton & Abrahams, 1972; Hamilton, 1974). The  $\delta R$  plot is a more powerful statistical descriptor than the traditional single-valued discrepancy index  $R = \sum |O_i - C_i| / \sum O_i$ .

**Order statistic:** When a sample of variate values are arranged in order of magnitude, these ordered values are known as order statistics.

**Parameter:** *Models* are formulated in terms of physical quantities called parameters, values for which are *estimated* from the observations. The true (unknown) value of a parameter is a constant, since it represents a physical fact independent of the observations. Any *estimate* of its value is a random variable. In the framework of *Bayesian* statistics, a parameter (not its estimate) is regarded as a random variable and the associated p.d.f. is taken to represent the scientist's belief in the value of the parameter.

**Population:** see *Random variable*.

**Precision:** The closeness of agreement between the values of a measurement or of an *estimate* obtained by applying a strictly identical experimental procedure several times. It is expressed numerically by a *standard deviation* or *variance*. The precision of a diffraction intensity is often inferred from only one or maybe two measurements by (1) invoking Poisson statistics for the count rates, and/or (2) using the experience gained from earlier diffraction experiments. Precise estimates are not necessarily accurate (Prince, 1985; Rollett, 1985).

**Probability density function (p.d.f.):** The function  $p(x)$  of the *random variable*  $x$ , such that the probability of finding  $x$  between  $a$  and  $b$  ( $a < b$ ) is given by

$$P(a \leq x \leq b) = \int_a^b p(x) dx. \quad (13)$$

The probability of finding  $x$  somewhere within the whole interval of variation is equal to unity.

Consider an ordered set, or  $n$  vector, of  $n$  *random variables*  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ . The **joint (or multivariate) p.d.f.** of  $\mathbf{x}$  is the function  $p_J(\mathbf{x})$ , such that

$$P(a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} p_J(\mathbf{x}) d^n \mathbf{x}. \quad (14)$$

The **marginal p.d.f.**  $p_M(x_i)$  of an element  $x_i$  of  $\mathbf{x}$  is the p.d.f. of that element irrespective of the values of any other elements:

$$p_M(x_i) = \int p_J(\mathbf{x}) d^{n-1} \mathbf{x}, \quad (15)$$

where the integration is over the full range of all elements **except**  $x_i$ . If  $p_J(\mathbf{x}, \mathbf{y})$  denotes the combined joint p.d.f. for the elements of two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , then the **conditional p.d.f.** for  $\mathbf{x}$  given  $\mathbf{y}$ ,  $p_C(\mathbf{x}|\mathbf{y})$ , is the joint p.d.f. of the elements of  $\mathbf{x}$  when the elements of  $\mathbf{y}$  are held at fixed particular values. It is related to the joint p.d.f. of  $\mathbf{x}$  and  $\mathbf{y}$  by

$$p_C(\mathbf{x}|\mathbf{y}) = p_J(\mathbf{x}, \mathbf{y}) / p_M(\mathbf{y}). \quad (16)$$

Therefore,  $p_J(\mathbf{x}, \mathbf{y}) = p_C(\mathbf{x}|\mathbf{y}) p_M(\mathbf{y}) = p_C(\mathbf{y}|\mathbf{x}) p_M(\mathbf{x})$  from which it follows that

$$p_C(\mathbf{x}|\mathbf{y}) = p_C(\mathbf{y}|\mathbf{x}) p_M(\mathbf{x}) / p_M(\mathbf{y}). \quad (17)$$

This last relation is known as **Bayes's theorem**. If  $p_C(\mathbf{x}|\mathbf{y}) = p_M(\mathbf{x})$  for all possible values of  $\mathbf{y}$ , the *random variables*  $\mathbf{x}$  and  $\mathbf{y}$  are statistically independent.

**Random error:** An error having a zero expected value. Particularly important in crystallography are statistical fluctuations in quantum counts. Other effects like irreproducible play in diffractometer settings, and short-term fluctuations of temperature, pressure, mains voltage (Abrahams, 1969) and humidity contribute, at least in part, to the random errors. Random errors can be reduced at the expense of increased measuring time. If the *model* and the method of refinement are perfect, errors of estimated parameters are also random (see *Bias*).

**Random variable:** The possible outcomes of a measurement. The set of all possible outcomes is a **population**. The set of outcomes of a finite number of repeated measurements is a **sample** drawn from the population.

**Repeatability:** The closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions, namely the same method of measurement, the same observer, the same measuring instrument, the same location, and repetition over a short period of time.

**Reproducibility:** The closeness of agreement between the results of measurements of the same measurand where the individual measurements are carried out changing conditions such as method of measurement, observer, measuring instrument, location, time.

**Residual:** A general term denoting a quantity remaining after some other quantity has been subtracted. It occurs in a variety of contexts. For example, the term residual has been used by crystallographers to mean the discrepancy index  $R$ . In a different sense, if the calculated value of a variable is subtracted from an observed value then the difference may be called a residual, although the more precisely defined term *deviate* is to be preferred. Some authors restrict it to mean only those deviates obtained from least squares at **convergence**, but terms such as 'deviates at convergence' or 'residuals at convergence' would avoid possible confusion.

**Sample:** see *Random variable*.

**Scaled deviance:** see *Deviance*

**Standard deviation:** The square root of the *variance* of a *probability density function*.

**Systematic error:** Contribution of the deficiencies of the *model* to the difference between an *estimate* and the true value of a quantity. A list of important systematic errors in measured X-ray intensities has been given by Abrahams (1969). It can be reduced by improving the model, but not by an investment in measuring time. In some cases, the presence of systematic errors can be inferred from the *goodness of fit*, from *normal probability plots* (Abrahams, 1974), from the distribution of scaled *deviates*  $(O_j - C_j)/\sigma_j$  as functions of  $O$ ,  $C$ ,  $(\sin \theta)/\lambda$ , diffractometer angles *etc.*, or by using the *Durbin-Watson* statistics (Flack, 1984, 1985) to reveal correlations. Some systematic errors (e.g. an incorrect value of the wavelength) cannot be detected by any statistical test. The amount of systematic error present in any *estimate* cannot be evaluated exactly (see *accuracy*). Note that, strictly speaking, observations are free of systematic error, in contrast to the *model* used for their interpretation (see *estimator*).

**Unbiased:** see *Estimator*.

**Uncertainty:** A colloquial term used to describe the lack of knowledge of the true value of a parameter that includes the effects of systematic error (model inadequacy), as well as lack of statistical precision.

**Variance:** The second moment about the mean (see *Moment*) of a *probability density function*.

**Variance-covariance matrix:** The  $n \times n$  matrix whose elements are the second *moments* about the mean of a joint *probability density function* of  $n$  *random variables* [see (10)].

**Weight:** Value used to express the relative importance of an observation (e.g. an intensity) with regard to the quantities to be deduced from the data set (e.g. atomic positional coordinates). Weights appear in

averaging, least squares, statistics (e.g. *goodness of fit*) and elsewhere. For a set of  $n$  data the weights are represented by an  $n \times n$  weights matrix  $W$ . Usually,  $W$  is taken to be diagonal. In averaging and least squares, minimum-variance estimates are obtained by  $W = V^{-1}$  where  $V$  is the *variance-covariance matrix* of the observations.

**Weighted deviance:** see *Deviance*.

## Refinement

Refinement is the process of adjusting the parameters of a *model* to find values most nearly compatible with the observations. From an approximate set of starting parameters for a model obtained by the methods of crystal structure determination, the parameter *estimates* are varied to obtain a best fit between the  $n$  observed quantities  $O_j$  and the corresponding calculated quantities  $C_j$ . In most crystallographic work, the  $O_j$  are diffraction intensities, but non-diffraction data may also be included. Concerning correction factors to obtain structure amplitudes  $|F|^2$  or  $|F|$ , we refer to the discussion of the term *model*.

The most popular refinement method in the physical sciences, and about the only one used in crystallography, is the method of least squares, which minimizes the weighted *deviance*  $D_w = \mathbf{d}^T W \mathbf{d}$ . Some statistical studies, however, suggest that least squares may not always be the best method. Tukey (1974) has asserted that chemists and physicists 'both make less of their data than they should and, too often, come to think better of their results than deserved'. While this may be true, no other methods have been shown to be convincing alternatives, except variants of least squares such as, for example, robust-resistant methods (see below). Among the other methods described in textbooks (Kendall & Stuart, 1979; Eadie, Drijard, James, Roos & Sadoulet, 1971), *maximum likelihood* takes a prominent place, but it is equivalent to least squares for a *normal* distribution of errors. The popularity of least squares is probably because the error distribution of the deviates  $O_j - C_j$  is not known (for the distribution of the net intensity see Wilson, 1980a). In the absence of this information, it is difficult to justify alternative approaches. In addition, most theoretical work assumes a normal distribution.

An authoritative discussion of the method of least squares is found in Prince (1982, 1985, 1989). Linear least squares is an **unbiased estimator** if the *weights* are independent of the observations, and if the *model* represents physical reality correctly for some set of values of the parameters. In particular, the **Gauss-Markov theorem** states that minimal *variances* of the *estimates* are obtained if the weight matrix  $W$  is chosen as the inverse of the *variance-covariance matrix*  $V$  of the joint p.d.f. of the observations. The inverse of the normal-equations matrix is an unbiased

estimate of the *variance-covariance matrix* of the model parameters if, and only if,  $W = V^{-1}$ . For any weighting scheme other than  $W = V^{-1}$ , the estimation of the *variance-covariance matrix* of the least-squares parameters should in principle be carried out using formulae given in Prince (1985, 1989) and in Rollett (1988), but this has never been tried in practice. The *goodness of fit* is expected to be 1.0. In crystallographic applications, the model is usually expressed by non-linear differentiable functions. The weighted *deviance* is minimized by iteratively linearizing the model functions with a Taylor expansion at approximate parameter values. The Gauss-Markov theorem applies then to the fit and parameter values at convergence insofar as the linearized functions are good approximations to the model functions in the vicinity of the minimum (see also Eadie *et al.*, 1971). In practice, the number of parameters to be estimated is considerable, and the use of off-diagonal terms in the weight matrix  $W$  is cumbersome in several respects. For this reason,  $W$  is usually chosen as a diagonal matrix. This practice is often justified by the presumption that intensity measurements are uncorrelated. However, off-diagonal terms in  $V$  may well arise if non-random errors such as absorption errors are present or if the measurements have been systematically altered.

The assumptions that the Gauss-Markov theorem is based upon are never realized in practice. In particular, the variances of the observations, whether they be derived from the spread of repeated or symmetry-equivalent observations about their average, or from Poisson statistics, are estimated from the observations themselves, by methods which are indeed a part of the model. The model is only an approximation to physical reality, and is often rather crude. The adjustable parameters may have no objective significance. Thus, harmonic and higher-order displacement formalisms only serve to parametrize the atomic p.d.f.'s. The values obtained by a refinement depend on the number of terms used in the expansion, and consequently only the total estimated p.d.f.'s may be physically meaningful. The true p.d.f.'s they approximate are not even guaranteed to possess the higher moments appearing in the Gram-Charlier and Edgeworth expansion formalisms (Johnson & Levy, 1974).

The assumed symmetry of the crystal structure is part of the model and is in many cases higher than the observed symmetry of the diffraction intensities. These are affected by anisotropic effects such as absorption and extinction for which corrections are only approximate. The symmetry-equivalent intensities, even if corrected for these effects, thus may belong to different *populations*. Consequently, their symmetry equivalence should not in principle be used as a criterion for averaging. Practical considerations of computation time and data storage

require, however, that some data reduction by way of averaging be carried out. Refinements, using a crystal structure, carried out by least squares on unaveraged and averaged data, are only identical under certain strict conditions\* rarely realized in practice. The significance of the *estimated standard deviations* of the average values of corrected intensities may be doubtful (see below for Bayesian interpretation).

It is not surprising that the *goodness-of-fit* values of most least-squares calculations on real data are only accidentally within an acceptable range near 1.0. The *variances* obtained for the *estimates* are thus of dubious significance. In fact, the results of independent determinations of the same structure may differ by much more, and hardly ever by less, than is allowed by statistical tests. The commonly applied procedure of multiplying e.s.d.'s by the *goodness-of-fit* value has no statistical basis. From the point of view of the Gauss-Markov theorem and frequentist statistics, this situation can only be improved by making the model more realistic.

The *Bayesian* interpretation of statistics incorporates prior knowledge or beliefs in the p.d.f. of the observations. In fact, measurements in all physical science are necessarily conditioned by what we expect to find *a priori*, and are thus not independent of the model. The *variance-covariance matrix* of the observations reflects the author's *estimation* not only of the variability of repeated measurements, but also of the effects of approximate or omitted corrections which are akin to model deficiencies. The deviations from the *average* of symmetry-equivalent  $|F|^2$  values may serve to estimate the anisotropy of such effects, and the standard practice of averaging can thus be justified. The confidence of the crystallographer in the model and in the *variance-covariance matrix* of the observations is based on an examination of the deviates and may be modified during refinement. The use of restraints also fits into the Bayesian philosophy, their estimated variances being chosen by the scientist's confidence in the restrained features of the model. The error estimate of the results is represented by an *estimated standard deviation*. Different scientists may arrive at different estimates, a degree of confidence being a subjective measure. The underlying model may be criticized and must, of course, be completely described.

---

\* These conditions are (a) the averaged data should consist of the weighted averages of the equivalent reflections calculated using the least-squares weights for the unaveraged data; (b) the least-squares weights for the averaged data should be the sum of those before averaging; (c) a weighted deviance term corresponding to the dispersion of the equivalent reflections about their weighted average should be added to the weighted deviance minimized for the averaged data. This deviance term does not alter the parameter values, but may affect the interpretation of the error estimates by altering the goodness of fit.

### Refinement on $I$ , $|F|^2$ or $|F|$ ?

There was strong disagreement among members of the Subcommittee over the question of whether 'observations' used in a refinement should be net integrated intensities  $I$ , values of  $|F|^2$  or values of  $|F|$ , or indeed whether it makes a difference. The critical factors in the transformation from peak and background scan intensities through net integrated intensities to  $|F|^2$  concern the application of correction terms associated with absorption, the Lorentz-polarization factor, thermal diffuse scattering *etc.*, whereas the change from  $|F|^2$  to  $|F|$  concerns the square-root function. The extraction of the square root is a non-linear operation that has the potential of introducing a *bias* proportional to the variance of  $|F|^2$  (Wilson, 1976*b*, 1979), with the additional problem of determining what to do for the very weak reflections where statistical fluctuations in the peak and background measurements may cause the net intensity to be negative. The widely used formula for the e.s.d. of  $|F|$ ,  $\sigma(|F|) = \sigma(|F|^2)/2|F|$ , may be appropriate for strong reflections but must be modified for weak reflections in order to prevent  $\sigma(|F|)$  from becoming infinite at  $|F| = 0$ . French & Wilson (1978) and Gonschorek (1985) have proposed methods for obtaining  $|F|$  and  $\sigma(|F|)$  from  $|F|^2$  and  $\sigma(|F|^2)$ .

The partial derivative of the calculated quantity  $C_j = |F|^n = (A^2 + B^2)^{n/2}$  with respect to the variable  $\nu_r$  is

$$\partial C_j / \partial \nu_r = n|F|^{n-2} \{A(\partial A / \partial \nu_r) + B(\partial B / \partial \nu_r)\}. \quad (18)$$

If the calculated structure factor is zero, *i.e.*  $A = B = 0$ , then  $\partial C_j / \partial \nu_r = 0$  for  $n \geq 2$ , and it is undefined for  $n < 2$ . The contributions of the  $j$ th observation to the normal-equations matrix and vector are  $w_j(\partial C_j / \partial \nu_r)(\partial C_j / \partial \nu_s)$  and  $w_j(O_j - C_j)(\partial C_j / \partial \nu_r)$ , respectively. If the weight is chosen according to  $w(|F_j|) = 4|F_j|^2 w(|F_j|^2)$ , then the contribution to the matrix is exactly the same, and the contribution to the vector nearly the same, for refinements on  $|F|$  and  $|F|^2$ . In many cases, omission of weak reflections has a negligible effect on the results (see recommendation 6), and the two kinds of refinement are then nearly identical. For this reason, the  $|F|$  versus  $|F|^2$  controversy is often considered to be irrelevant.

**The arguments in favor of refinement on  $|F|$**  are based on a mathematical analysis by Prince & Nicholson (1985). They observe that different reflections have different **leverage** which is a quantity that measures the influence of an individual reflection on the fit. It is proportional to the contributions to the matrix and vector described above. Because  $\partial|F|^2/\partial\nu_r$  is small if  $|F|^2$  is small, weak reflections have little leverage in refinements on  $|F|^2$ , even if some of the derivatives  $\partial F/\partial\nu_r$  are substantial. Including them in a refinement on  $|F|^2$  can do no harm, but no good either.

In a refinement on  $|F|$ , care must be taken to give non-zero weights to the weak reflections, as mentioned before. Regarding the discontinuity of the partial derivatives of  $|F_c|$  at  $F_c = 0$ , Prince & Nicholson (1985) argue that the practice of assigning the phase of  $F_c$  to  $|F_o|$  in effect makes the 'observation' a Bayesian prior estimate of  $F$  (not  $|F|$ ), and that the equivalent value for an unobserved reflection is  $F = 0$  with a finite variance based on the threshold of observability. The partial derivatives of  $F$  with respect to the model parameters are continuous everywhere. These reflections, even though weak, may be sensitive to some parameters and do not lose their leverage. Their inclusion can improve the precision of parameter estimates.

**The arguments in favor of refinement on  $|F|^2$**  are based on the *bias* introduced by extracting the square root (Wilson, 1976*b*, 1979), and on the undesirable discontinuity of the partial derivatives of  $|F|$  at  $F = 0$ . In principle, refinements should be on quantities as near as possible to the actual observations, and all non-linearities should be part of the model. It is improper to exclude the weak reflections, where the bias may be a substantial fraction of the  $|F|$  value, by introducing an arbitrary cut-off at a minimum net intensity, be it zero or some positive value, because this results in a systematic inclusion of reflections with positive fluctuations and exclusion of reflections with negative fluctuations, and thus in a biased data set. Bias is therefore avoided by using **all** reflections in a refinement on  $|F|^2$ . Hirshfeld & Rabinovich (1973) recommend inclusion of negative net intensities with their negative values. The discontinuity of the partial derivatives of  $|F|$  is akin to the crystallographic phase problem. The model contains little information on the phase of  $F$  if the calculated structure factor is small, and none at all at  $F_c = 0$ . For these reflections, we thus do not have a good Bayesian prior estimate of  $F$ , hence their variances should be large and their small leverage in a refinement on  $|F|^2$  appears to be justified. The leverage is not given by the 'observed', but by the **calculated** value  $|F_c|^2$ . Thus, a weak reflection may have a non-negligible effect if  $|F_c|^2$  is considerably larger than  $|F_o|^2$ . If, on the other hand,  $|F_c|^2 = 0$  while  $|F_o|^2$  is larger than maybe 5 e.s.d.'s, the structure presents an unsolved phase problem. Apart from a possible bias when refining on  $|F|$ , the main difference between refinements on  $|F|$  and  $|F|^2$  is equivalent to an up-weighting of weak reflections if  $\sigma(|F|)$  is kept finite for  $|F| = 0$ . The same effect could be achieved by explicitly up-weighting the weak reflections. In some structures, weak reflections are of prime importance. Crystallographers working on such structures are urged to give serious attention to these opposing arguments (see recommendation 6).

**The arguments in favor of refinement on  $I$**  are an extension of those for refinement on  $|F|^2$ . It has



become customary to regard the relationships between peak and background measurements, net intensity and  $|F|^2$  as being linear transformations with **constant** factors. In this way, the measurement error (or the confidence in *a priori* estimates) of crystal dimensions, polarization ratio, changes in reference intensity for radiation damaged or decomposing crystals amongst other systematic effects are not taken into account. The transformation from peak and background measurements to net intensity values leads to the paradox of 'negative as-measured' intensities. As systematic error is a major source of trouble, some members of the Subcommittee propose investigation of methods permitting refinement of crystal structures on **all** observations, *i.e.* measured intensities as well as crystal dimensions, polarization ratio, crystal decomposition curves *etc.* They propose inclusion in the model of additional refineable parameters (*e.g.* crystal dimensions) and introduction of the corresponding observations as restraints. The advantages are clearly an improved modeling of the structure and more realistic error estimates on atomic parameters.

### Defects in the model

It is evident that the results of a refinement can be improved by identifying the presence and origin of systematic error, and by improving the model to satisfy the assumptions of frequentist statistics (Gauss-Markov theorem) more fully. The almost universally observed deviation of the goodness-of-fit value from unity indicates the presence of defects in the model and/or the *variance-covariance matrix* of the observations (see also Wilson, 1980*b*). A *normal probability plot* gives more detailed information on the presence of systematic error.

Identifying the origin of systematic error and improving the model is far more difficult. Beu and his collaborators (Beu, Musil & Whitney, 1962, 1963; Beu & Whitney, 1967; see also Mitra, Ahmed & Das Gupta, 1985; Mandel, 1980) achieved a major improvement in the precise and accurate determination of lattice parameters by careful tests and corrections for remaining systematic errors based on *maximum likelihood*. They assumed a *normal* distribution of errors, and their refinement technique was therefore equivalent to least squares (Wilson, 1980*b*). The modeling of electron density distributions with aspherical-atom formalisms (Stewart, 1976; Hirshfeld, 1977), and of atomic thermal-displacement probability density functions with anharmonic contributions (Johnson & Levy, 1974; Kuhs, 1983) has shown considerable success, but has not had a significant impact on standard crystal-structure determination, which in its present form answers the needs of chemical crystallographers and will continue to be an increasingly automated analytical technique.

A thorough study of intensity measurement and data-reduction procedures might indicate more generally applicable improvements.

### Weighting schemes

A very commonly used criterion for the choice of weights is that the variance of the derived estimates be minimal, although other weighting schemes are permissible. The minimum-variance criterion implies that  $W = V^{-1}$ . Regardless of the choice of weights, proper calculation of *estimated standard deviations* and of the *goodness of fit* in intensity averaging and in least squares will require an estimate of  $V$ , the *variance-covariance matrix* of the observations. The frequentist and *Bayesian* approaches to statistics lead to different interpretations of  $V$  (see section on *Refinement*). For the one,  $V$  is concerned only with the random fluctuations of the measurements whereas, for the other,  $V$  also incorporates the scientist's belief in the model with which the observations are being analysed.

### Weights of averaged intensities

Minimum-variance weights can at best be based on *estimated standard deviations* of intensities. In many cases,  $n$  symmetry-equivalent corrected net intensities  $I_i = c_i O_i$  will be averaged, where the  $c_i$  will contain at least the scan speeds and absorption corrections. Weights used in averaging should **not** be based on the counting statistics of the individual observations whose estimated variances are biased and result in larger weights for accidentally low intensities, and lower weights for accidentally high intensities. Owing to the symmetry postulated while averaging, the expected values of all corrected intensities  $I_i$ ,  $E(I_i)$ , must be the same quantity  $\mu_I$ , and the expected values of the observed intensities  $O_i$  are thus

$$E(O_i) = \mu_I / c_i. \quad (19)$$

An often used expression for the variance of  $O_i$  (Abrahams, 1974) is

$$\begin{aligned} \sigma^2(O_i) &= E(O_i) + b_i + kE(O_i)^2 \\ &= \mu_I / c_i + b_i + k\mu_I^2 / c_i^2, \end{aligned} \quad (20)$$

where  $b_i$  is the contribution of the background estimated from peak and background count rates, and  $kE(O_i)^2$  is the contribution from known and unknown sources of random error. Minimum-variance weights  $\sigma^{-2}(I_i) = \{c_i \sigma(O_i)\}^{-2}$  needed for the calculation of the average intensity can be obtained by iteratively replacing  $\mu_I$  by an approximate average value  $\bar{I}$ . In a first step,  $b_i$  and  $k$  may be neglected and the average intensity then becomes,

$$\bar{I} = \sum (I_i / c_i) / \sum (1 / c_i) = \sum O_i / \sum (1 / c_i). \quad (21)$$

Estimates of the factor  $k$  in (20) can be obtained from

the variations of periodically measured check reflections. Differences among symmetry-equivalent intensities can (and indeed should) also be used to estimate  $k$ , the value of which then includes at least part of an anisotropic *systematic error*. A possible procedure consists in adjusting  $k$  to obtain comparable values for the variance of  $\bar{I}$  estimated from the  $\sigma^2(I_i)$  derived from (20), and estimated from the spread of the  $I_i$  around the average value  $\bar{I}$ :

$$\begin{aligned}\sigma^2(\bar{I}) &\approx 1/\sum^n \sigma^{-2}(I_i) \\ &\approx \left[ \sum^n \sigma^{-2}(I_i) \{I_i - \bar{I}\}^2 \right] / \left[ (n-1) \sum^n \sigma^{-2}(I_i) \right]; \\ \sigma(I_i) &= c_i \sigma(O_i); \quad n \geq 2.\end{aligned}\quad (22)$$

The average goodness of fit of all symmetry-equivalent sets is then near 1.0. If the number  $n$  of symmetry-equivalent observations is large, the larger of the two estimates of  $\sigma(\bar{I})$  obtained with some approximation for  $\sigma^2(I_i)$ , their average, or the value from the spread alone may be used. Alternative and more convenient procedures, variants and extensions may be proposed, or have already been implemented (see also Blessing, 1987).

#### Weights based on the model

The theory of least squares shows that expected values of parameter estimates are not affected by the choice of weights, provided that the model is free of *systematic error* (i.e. the *estimator* is unbiased), and the weights are not functions of the *deviates*  $d_j = O_j - C_j$  (Prince & Nicholson, 1985; Prince, 1985, 1989). On the other hand, weights based on  $d_j$ ,  $O_j$  or  $C_j$  may result in a bias. As remarked before, variances of the observations are usually estimated from the observations themselves. In particular, the contribution of counting statistics to the variance of an intensity is proportional to the intensity itself. An accidentally low intensity will have a lower variance than an accidentally high intensity, showing that this estimate of the variance is indeed biased. The corresponding weights may result in bias in the parameter estimates. Wilson (1976*b*) investigated the effect by assuming that the weight of a *deviate* is an arbitrary function of the corresponding measured and/or calculated quantity,  $w_j = w_j(O_j, C_j)$ . He showed that in case of refinement of one parameter, the use of the weighted mean  $w_j\{O_j + 2C_j\}/3$  removes this bias to the order of the mean-square statistical fluctuation of the measurement. Weights from (20) and (22) may be functions of all intensities, and the corresponding bias is then more difficult to evaluate.

The omission of observations is equivalent to assigning zero weights. Again, a bias may be introduced if the criterion for omission is based on  $d_j$ ,  $O_j$  or  $C_j$ . Thus, if weak reflections are systematically

omitted, those with an accidentally low intensity will be preferentially discarded, and those with an accidentally high intensity will have a better chance of being retained. An example of such a bias has been described by Seiler, Schweizer & Dunitz (1984). Omission of observations according to  $(\sin \theta)/\lambda$  or reflection number should *not* result in a bias.

Weight-modification schemes designed to lessen the effects of *systematic errors* are necessarily based on the *deviates*  $d_j$ . An example is found in Wang & Robertson (1985). They estimate the variances of the structure amplitudes from the distribution of the *weighted deviates*. The resulting weights are intended to represent all kinds of *random errors* from different sources, but include also contributions of *systematic errors*. The **robust/resistant refinement techniques** are another example. The results of a standard least-squares refinement may be considerably influenced by 'outliers', observations that seem so discrepant with calculated values that a blunder of the investigator or a malfunctioning of the equipment is suspected. The naive approach to such observations is to discard them by setting the weights to zero whenever the discrepancy exceeds a certain limit. The robust/resistant methods avoid the discontinuity of this procedure by decreasing the weights gradually with increasing disagreement (Prince, 1982; Nicholson, Prince, Buchanan & Tucker, 1982).

#### Restraints

Restraints or soft constraints are relations between parameters of the model that are treated formally in the same way as observations (Waser, 1963; Hendrickson & Konnert, 1980; Hendrickson, 1985). They may be used to specify bond lengths, angles, planarity of molecules, relations between displacement parameters, and for numerous other purposes. The variances (and covariances) attributed to such pseudo-observations are chosen according to the scientist's beliefs regarding their validity, in agreement with the *Bayesian* interpretation of statistics. E.s.d.'s of the parameter estimates are, of course, dependent on these variances. It may be difficult to judge whether they are appropriate if they are recorded in a structure data file that omits all information regarding the restraints.

#### Enhancing particular features

A minority of Subcommittee members believes in the merit of weighting schemes chosen to enhance particular features of the structure. These weights then reflect the scientist's aims rather than the precision of the measurements. A more demanding method would consist in altering the experimental design to correspond more closely to the objectives of the study. Prince & Nicholson (1985) address the problem of finding the reflections that should be

measured more precisely in order to enhance particular features.

Obtaining an unweighted least-squares fit between observed and calculated  $|F|^2$  is equivalent to obtaining an unweighted least-squares fit between observed and calculated Patterson functions. Similarly, obtaining an unweighted least-squares fit on the  $|F|$ 's is almost equivalent to obtaining an unweighted least-squares fit between observed and calculated electron densities (Wilson, 1976a).

Special weighting schemes permit more accurate fitting of the electron densities at the atomic sites. Thus, Cochran (1948, 1951) showed that refinement on the  $|F|$ 's with weights inversely proportional to the reciprocal of the atomic scattering factor of an atom gives the coordinates of the maximum of its electron density as represented by a Fourier series. In the presence of several types of atoms, weighting with the reciprocal of a mean scattering factor is probably a satisfactory approximation. Similarly, the weighting scheme of Dunitz & Seiler (1973) designed to determine more accurately the coordinates of the atomic centers also emphasizes high-order reflections.

Bernardinelli & Flack (1985) have designed weights that enhance the sensitivity of a refinement to the centrosymmetric or antisymmetric parts of the electron density, and also to the absolute structure. They are useful in resolving ambiguities between centro- and non-centrosymmetric structures and mainly affect weak reflections, the importance of which in such problems has been pointed out by Marsh (1981, 1986). Weak reflections are also critically important in the refinement of superstructures, and weighting schemes are easily designed to enhance their contribution to the normal equations.

### Recommendations

It is not realistic to expect that a statistical procedure can prevent or identify careless work. Thus, published lattice constants with reasonably small e.s.d.'s may be grossly in error, owing to a variety of causes (*cf.* Parrish, 1960). Only independent redeterminations may show this. **Thoughtless use of established procedures in widely distributed software may be as harmful as the natural tendency of most people to prefer results in agreement with preconceived ideas.** Note, however, that preconceived ideas are an ingredient of Bayesian statistics. Since the precision may be evaluated with greater confidence than the accuracy, it is not surprising that the results of independent determinations of the same structure may differ by much more, and hardly ever by less than is allowed by statistical tests. Inter-determination *goodness-of-fit* values can give an indication of the average discrepancies to be expected (Taylor & Kennard, 1986). The following recommendations are intended to pro-

duce more meaningful results from structure determinations:

1. All reflections to be used in the refinement process should be measured more than once, preferably in more than one symmetry-equivalent position. A fixed length of time allotted to an experiment is better used for rapid measurement of symmetry-equivalent reflections or for integrated intensity measurements made at different values of the azimuthal angle than for more precise measurements limited to a single independent portion of reciprocal space (Ibers, 1967; Hamor, Steinfink & Willis, 1985). (It should be noted that additional measurements on a different crystal using perhaps a different radiation can also be very revealing concerning the accuracy of refined parameters.)

2. If possible, the shape of the sample crystal should routinely be measured and absorption corrections applied.

3. Estimated variances of the intensities should not be based on counting statistics alone. They should at least take into account the variation of *several* periodically measured check reflections, which indicate a possible minimal value of  $k$  in (20). In the case of averaged intensities, differences among absorption-corrected symmetry-equivalent data and measurements at different values of the azimuthal angle can be used in the estimation. It is an act of faith to assume that these differences are equivalent to random fluctuations, but the uncertainty can be **no less than** that inferred from this assumption.

4. Measured and calculated values of  $I$ ,  $|F|^2$  or  $|F|$ , with e.s.d.'s for **all** measured values, should be provided to referees. Any serious discrepancies between measured and calculated values should be noted and commented upon in the text. Calculated values need not be deposited.

5. E.s.d.'s of derived quantities like angles and bond lengths should be calculated using the full *variance-covariance matrix* of the refined parameters. The programming effort and additional calculation time necessary for a full variance-covariance propagation-of-error analysis are very well rewarded by greatly improved estimated standard deviations. In some cases, *e.g.* in the presence of pseudo-symmetry, the use of the full variance-covariance matrix must even be considered mandatory.

6. All intensities or structure amplitudes should be included in the refinement. Omitting weak reflections often has little effect on the results (Seiler, Schweizer & Dunitz, 1984; Wang, Barton & Robertson, 1986). However, weak reflections may contain important information: they may be vitally important when a choice is to be made between a centrosymmetric and a non-centrosymmetric model (Marsh, 1981, 1986), and when refining superstructures. Intensities measured as negative may be left negative in the refinement (Hirshfeld & Rabinovich, 1973). Setting

them to zero (**after** averaging) is advocated with the argument that even a perfect model cannot reproduce negative intensities. They may also be set to positive values by a *Bayesian* procedure (French & Wilson, 1978). For further discussion of weak reflections see the section *Refinement on I,  $|F|^2$  or  $|F|$ ?*

7. It is of utmost importance to resolve any space-group ambiguities, and in particular to ascertain the presence or absence of a center of symmetry. The methods for doing this fall into two classes:

(i) *Modern statistical tests operating on diffraction data*

Tests based on the Wilson (1949) statistics using complete sets of intensity data perform remarkably well for structures containing a large number of not too dissimilar atoms which occupy general positions in the asymmetric unit of the space group. An additional requirement is the absence of non-crystallographic symmetry.

Approximate methods which may cope with the presence of outstandingly heavy atoms were brought to a form applicable to all space groups by Shmueli & Wilson (1981) and Shmueli & Kaldor (1981, 1983). In practice, these methods may fail for low-symmetry structures with extreme atomic heterogeneities.

Exact probability distributions of structure amplitudes which are formulated as Fourier series and can be computed to any precision are now available for low-symmetry space groups (Shmueli, Weiss, Kiefer & Wilson 1984; Shmueli & Weiss, 1987). They allow for hypercentric distributions (Shmueli, Weiss & Kiefer, 1985), and for heavy scatterers in special positions (Shmueli & Weiss, 1988). These methods account correctly for any atomic heterogeneity.

(ii) *Measurement of symmetry-dependent physical properties*

These properties include crystal morphology, etch figures, optical activity, pyroelectricity and piezoelectricity (*International Tables for Crystallography*, 1983), which generally require the availability of single crystals with no linear dimension less than about 2 mm. The most powerful and discriminating test rests upon detection of the generation of second harmonics, see for example Dougherty & Kurtz (1976). This method requires only the availability of a microcrystalline sample. All these methods may reveal the absence, but not the presence, of a center of symmetry.

8. The technique of multiplying the terms of the estimated *variance-covariance matrix* of the model parameters by the square of the value of the *goodness of fit* involves a highly questionable assumption. In general, variances should not be adjusted to modify the goodness of fit. The goodness of fit should be reported in publications.

9. Reliability indices like *R*, *wR* and *goodness of fit S* give a global measure of fit. They are not well suited for testing certain properties of the structure, such as polarity, absolute configuration or the pres-

ence of a center of symmetry. An improved global fit obtained by modifying one of these properties represents also all concomitant changes of the refined parameters and may not indicate the correctness of the modification. Rather, the property should be represented by a single refinable parameter, whose refined value and e.s.d. is much more indicative. This has been successfully done for the determination of absolute configuration and polarity (Rogers, 1981; Flack, 1983).

It is a pleasure to thank our colleagues associated with the IUCr and particularly Dr D. J. Finney of the International Statistical Institute for clarifying comments on an earlier draft.

## References

- ABRAHAMS, S. C. (1969). *Acta Cryst.* **A25**, 165–173.
- ABRAHAMS, S. C. (1974). *Acta Cryst.* **B30**, 261–268.
- ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* **A27**, 157–165.
- BAYES, T. (1763). *Philos. Trans. R. Soc. London*, **53**, 370–418.
- BERNARDINELLI, G. & FLACK, H. D. (1985). *Acta Cryst.* **A41**, 500–511.
- BEU, K. E., MUSIL, F. J. & WHITNEY, D. R. (1962). *Acta Cryst.* **15**, 1292–1301.
- BEU, K. E., MUSIL, F. J. & WHITNEY, D. R. (1963). *Acta Cryst.* **16**, 1241–1242.
- BEU, K. E. & WHITNEY, D. R. (1967). *Acta Cryst.* **22**, 932–933.
- BLESSING, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.
- BOX, G. E. P. & TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.
- COCHRAN, W. (1948). *Acta Cryst.* **1**, 138–142.
- COCHRAN, W. (1951). *Acta Cryst.* **4**, 408–411.
- DOUGHERTY, J. P. & KURTZ, S. K. (1976). *J. Appl. Cryst.* **9**, 145–158.
- DUNITZ, J. D. & SEILER, P. (1973). *Acta Cryst.* **B29**, 589–595.
- DURBIN, J. & WATSON, G. S. (1950). *Biometrika*, **37**, 409–428.
- DURBIN, J. & WATSON, G. S. (1951). *Biometrika*, **38**, 159–178.
- DURBIN, J. & WATSON, G. S. (1971). *Biometrika*, **58**, 1–19.
- EADIE, W. T., DRIJARD, D., JAMES, F. E., ROOS, M. & SADOULET, B. (1971). *Statistical Methods in Experimental Physics*. Amsterdam: North-Holland.
- FLACK, H. D. (1983). *Acta Cryst.* **A39**, 876–881.
- FLACK, H. D. (1984). In *Methods and Applications in Crystallographic Computing*, edited by S. R. HALL & T. ASHIDA, pp. 41–55. Oxford: Clarendon Press.
- FLACK, H. D. (1985). In *Crystallographic Computing 3: Data Collection, Structure Determination, Proteins, and Databases*, edited by G. M. SHELDRICK, C. KRÜGER & R. GODDARD, pp. 18–27. Oxford: Clarendon Press.
- FLACK, H. D., VINCENT, M. G. & VINCENT, J. A. (1980). *Acta Cryst.* **A36**, 495–496.
- FRENCH, S. (1978). *Acta Cryst.* **A34**, 728–738.
- FRENCH, S. & OATLEY, S. (1982). In *Crystallographic Statistics*, edited by S. RAMASESHAN, M. F. RICHARDSON & A. J. C. WILSON, pp. 19–51. Bangalore: Indian Academy of Sciences.
- FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* **A34**, 517–525.
- GONSCHOREK, W. (1985). *Acta Cryst.* **A41**, 189–195.
- HAMILTON, W. C. (1964). *Statistics in Physical Science*. New York: Ronald Press.
- HAMILTON, W. C. (1974). *International Tables for X-ray Crystallography*, Vol. IV, pp. 293–294. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
- HAMILTON, W. C. & ABRAHAMS, S. C. (1972). *Acta Cryst.* **A28**, 215–218.

- HAMOR, T. A., STEINFINK, H. & WILLIS, B. T. M. (1985). *Acta Cryst.* **C41**, 301–303.
- HENDRICKSON, W. A. (1985). In *Methods in Enzymology*, Vol. 115. *Diffraction Methods for Biological Macromolecules, Part B*, edited by H. W. WYKOFF, C. H. W. HIRS & S. N. TIMASHEFF, pp. 252–270. New York: Academic Press.
- HENDRICKSON, W. A. & KONNERT, J. H. (1980). In *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & D. VENKATESAN, pp. 13.01–13.26. Bangalore: Indian Academy of Sciences.
- HILL, R. J. & FLACK, H. D. (1987). *J. Appl. Cryst.* **20**, 356–361.
- HIRSHFELD, F. L. (1977). *Isr. J. Chem.* **16**, 226–229.
- HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* **A29**, 510–513.
- IBERS, J. A. (1967). *Acta Cryst.* **22**, 604–605.
- International Tables for Crystallography* (1983). Vol. A, pp. 780–785. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)
- JAYNES, E. T. (1983). In *Papers on Probability, Statistics and Statistical Physics*, edited by R. O. ROSENKRANTZ, pp. 116–130. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)
- JOHNSON, C. K. & LEVY, H. A. (1974). *International Tables for X-ray Crystallography*, Vol. IV, pp. 314–319. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
- KENDALL, M. G. & STUART, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, 4th ed. London: Griffin.
- KUHS, W. F. (1983). *Acta Cryst.* **A39**, 148–158.
- MANDEL, J. (1980). In *Accuracy in Powder Diffraction*, edited by S. BLOCK & C. HUBBARD, pp. 353–359. Washington, DC: US Government Printing Office.
- MARSH, R. E. (1981). *Acta Cryst.* **B37**, 1985–1988.
- MARSH, R. E. (1986). *Acta Cryst.* **B42**, 193–198.
- MITRA, G. B., AHMED, R. & DAS GUPTA, P. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 151–181. Guilderland, NY: Adenine Press.
- MORINO, Y., KUCHITSU, K. & MURATA, Y. (1965). *Acta Cryst.* **18**, 549–557.
- MURATA, Y. & MORINO, Y. (1966). *Acta Cryst.* **20**, 605–609.
- NICHOLSON, W. L., PRINCE, E., BUCHANAN, J. & TUCKER, P. (1982). In *Crystallographic Statistics*, edited by S. RAMASESHAN, M. F. RICHARDSON & A. J. C. WILSON, pp. 229–263. Bangalore: Indian Academy of Sciences.
- PARRISH, W. (1960). *Acta Cryst.* **13**, 838–850.
- PRINCE, E. (1982). *Mathematical Techniques in Crystallography and Materials Science*. New York: Springer.
- PRINCE, E. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 95–103, 107. Guilderland, NY: Adenine Press.
- PRINCE, E. (1989). *International Tables for Crystallography*, Vol. C. In preparation.
- PRINCE, E. & NICHOLSON, W. L. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 183–195. Guilderland, NY: Adenine Press.
- ROGERS, D. (1981). *Acta Cryst.* **A37**, 734–741.
- ROLLETT, J. S. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 105–107. Guilderland, NY: Adenine Press.
- ROLLETT, J. S. (1988). *Crystallographic Computing 4: New Techniques and New Technologies*, edited by N. W. ISAACS & M. R. TAYLOR, pp. 149–166. Oxford Univ. Press/International Union of Crystallography.
- SEILER, P., SCHWEIZER, W. B. & DUNITZ, J. D. (1984). *Acta Cryst.* **B40**, 319–327.
- SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* **A37**, 76–80.
- SHMUELI, U. & KALDOR, U. (1983). *Acta Cryst.* **A39**, 615–621.
- SHMUELI, U. & WEISS, G. H. (1987). *Acta Cryst.* **A43**, 93–98.
- SHMUELI, U. & WEISS, G. H. (1988). *Acta Cryst.* **A44**, 413–417.
- SHMUELI, U., WEISS, G. H. & KIEFER, J. E. (1985). *Acta Cryst.* **A41**, 55–59.
- SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* **A40**, 651–660.
- SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* **A37**, 342–353.
- STEWART, R. F. (1976). *Acta Cryst.* **A32**, 565–574.
- TAYLOR, R. & KENNARD, O. (1986). *Acta Cryst.* **B42**, 112–120.
- THEIL, H. & NAGAR, A. L. (1961). *J. Am. Stat. Assoc.* **56**, 793–806.
- TUKEY, J. W. (1974). In *Critical Evaluation of Chemical and Physical Structural Information*, edited by D. R. LIDE & A. P. MARTIN, pp. 3–14, 48–58. Washington, DC: National Academy of Sciences.
- WANG, H., BARTON, R. J. & ROBERTSON, B. E. (1986). *Program Abstracts. Am. Crystallog. Assoc. Annu. Meet.*, Hamilton, Ontario, Ser. 2, Vol. 14, p. 23. American Crystallographic Association.
- WANG, H. & ROBERTSON, B. E. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 125–135. Guilderland, NY: Adenine Press.
- WASER, J. (1963). *Acta Cryst.* **16**, 1091–1094.
- WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
- WILSON, A. J. C. (1976a). *Acta Cryst.* **A32**, 781–783.
- WILSON, A. J. C. (1976b). *Acta Cryst.* **A32**, 994–996.
- WILSON, A. J. C. (1979). *Acta Cryst.* **A35**, 122–130.
- WILSON, A. J. C. (1980a). *Acta Cryst.* **A36**, 929–936.
- WILSON, A. J. C. (1980b). *Acta Cryst.* **A36**, 937–944.

*Acta Cryst.* (1989). **A45**, 75–85

## Two-Dimensional Intensity Distributions Obtained with Crystal-Monochromated X-radiation and a Small Specimen Crystal

BY ANDREW W. STEVENSON

*Division of Materials Science and Technology, CSIRO, Locked Bag 33, Clayton, Victoria, Australia 3168*

(Received 29 February 1988; accepted 21 July 1988)

### Abstract

Two-dimensional ( $\Delta\omega$ ,  $\Delta 2\theta$ ) intensity distributions in the plane of diffraction are calculated for the case of crystal-monochromated X-radiation incident on a

small specimen. The calculations are based on ray tracing and take into account the mosaic spread of the monochromator, depth of penetration into the monochromator, source emissivity distribution, wavelength distribution, broadening due to the