International Tables for Crystallography (2006). Vol. C, Section 7.5.3, pp. 666–667.

7.5. Statistical fluctuations

By A. J. C. Wilson

7.5.1. Distributions of intensities of diffraction

Intensities of diffraction have two distinct probability distributions: (1) the *a priori* probability that an arbitrarily chosen reflection of a particular substance will have a particular 'true' intensity (R in the notation used below), and (2) the probability that a 'true' R will have an observed value R_o. The distributions of the first type depend on the symmetry and composition of the material, and are treated in Chapter 2.1 of Volume B of *International Tables for Crystallography* (Shmueli, 1993). The distributions of fluctuations of the second type, variations of the values R_o observed for a particular reflection, are treated here.

The crude counts or counting rates are rarely used directly in crystallographic calculations; they are subjected to some form of data processing to provide 'intensities of reflection' I_o in a form suitable for the determination of crystal structures, electron densities, line profiles for the study of defects, *etc*. The resulting intensity for the jth reflection, $I_{o,j}$, will be directly proportional to the corresponding $R_{o,j}$; when no confusion can arise, one of the subscripts will be omitted. The value of the proportionality factor c_j in

$$I_{o,j} = c_j R_{o,j} \qquad (7.5.1.1)$$

will be different for different reflections, since they will occur at different Bragg angles, have different absorption corrections, *etc*.

7.5.2. Counting modes

Whatever the radiation, in both single-crystal and powder diffractometry, the integrated intensity of a reflection is obtained as a difference between a counting rate averaged over a volume of reciprocal space intended to include the reflected intensity and a counting rate averaged over a neighbouring volume of reciprocal space intended to include only background. If these intentions are not effectively realized, there will be a systematic error in the measured intensity, but in any case there will be statistical fluctuations in the counting rates. The two basic modes (Parrish, 1956) are fixed-time counting and fixed-count timing. In the first, counts are accumulated for a pre-determined time interval, and the variance of the observed counting rate is proportional to the true (mean) counting rate. In the second, on the other hand, the counting is continued until a pre-determined number of counts is reached, and the variance of the observed counting rate is proportional to the square of the true counting rate. Put otherwise, the relative error in the intensities goes down inversely as the square root of the intensity for fixed-time counting, whereas it is independent of the intensity for fixed-count timing. Each mode has advantages, depending on the purpose of the measurements, and numerous modifications and compromises have been proposed in order to increase the efficiency of the use of the available time. References to some of the many papers are given in Section 7.5.7.

In principle, probability distributions can be determined for any postulated counting mode. In practice, they become complicated for all but the simplest modes; this is true even for the single measurement of the total counting rate or the background counting rate, but is even more pronounced for the distribution of their difference (the reflection-only rate). For most crystallographic purposes, however, it is only necessary to know the mean (to correct for bias, if present) and the variance (for the estimation of weights in refinement processes, see Part 8) of the distribution function.

7.5.3. Fixed-time counting

In the absence of disturbing influences (mains-voltage fluctuations, unrectified or unsmoothed high-tension supplies, 'dead time' of the counter or counter circuits, *etc*.), the number of counts recorded during the predetermined time interval used in the fixed-time mode will fluctuate in accordance with the Poisson probability distribution. If the 'true' number of counts to be expected in the interval is N, the probability that the observed number will be N_o is given by

$$p(N_o) = \exp(-N) N^{N_o}/N_o!, \qquad (7.5.3.1)$$

where all quantities appearing are necessarily non-negative. Both the mean and the variance of N_o are N. If the 'true' number of counts to be expected when the diffractometer is set to receive a reflection is T, and the 'true' number when it is set to receive the immediate background is B, the 'true' intensity of the reflection is

$$R = T - B, \qquad (7.5.3.2)$$

provided that the time interval used for the reflection is equal to the time interval used for the background. In practice, the 'observed' values of T_o and B_o fluctuate with probabilities given by (7.5.3.1) with T or B replacing N, so that the observed value,

$$R_o = T_o - B_o, \qquad (7.5.3.3)$$

will also fluctuate, and can, occasionally, take on negative values. The treatment of 'measured-as-negative' intensities is discussed in Section 7.5.6.

Although the sum of two Poisson-distributed variables is also Poisson, the difference is not, and the probability of R_o given by (7.5.3.3) has been shown to be (Skellam, 1946; Wilson, 1978)

$$p(R_o) = \exp\{-(B+T)\}(T/B)^{R_o/2} I_{|R_o|}[2(BT)^{1/2}], \qquad (7.5.3.4)$$

where I_n is the hyperbolic Bessel function of the first kind. The mean and variance of R_o are $T - B$ and $T + B$, respectively. If the times used for reflection and background are not equal, but are in the ratio of $k:1$, the mean and variance of

$$R_o = T_o - kB_o \qquad (7.5.3.5)$$

are $T - kB$ and $T + k^2 B$, respectively. The distribution of R_o then involves a generalized Bessel function* depending on k:

$$p(R_o) = \exp\{-(B+T)\}(T^{2-k}/B)^{R_o/2} I^k_{|R_o|}[2(BT^k)^{1/2}]; \qquad (7.5.3.6)$$

for the properties of the generalized Bessel function, see Wright (1933) and Olkha & Rathie (1971). The probability distributions (7.5.3.4) and (7.5.3.6) seem to be very close to a normal distribution with the same mean and variance for R_o near R, but the probability of moderate deviations from R is less than normal and for large deviations is greater than normal.

In applications, it is usual to work with intensities expressed as counting rates, rather than as numbers of counts, even when fixed-time counting is used. The total intensity τ expressed as a counting rate is

* The term 'generalized Bessel function' seems to have no unique meaning in mathematics. The functions given that name by Paciorek & Chapuis (1994) belong to a different group.

$$\tau = T/t, \tag{7.5.3.7}$$

where t is the time devoted to the measurement, and the variance of the counting rate is

$$\sigma^2(\tau) = T/t^2 = \tau/t. \tag{7.5.3.8}$$

Similar expressions apply for the background, with B for the count, b for the time, and β for the counting rate. For the reflection count, the corresponding expressions are

$$\rho = T/t - B/b, \tag{7.5.3.9}$$

$$\sigma^2(\rho) = \tau/t + \beta/b. \tag{7.5.3.10}$$

To avoid confusion, upper-case italic letters are used for *numbers* of counts, lower-case italic for counting *times*, and the corresponding lower-case Greek letters for the corresponding counting *rates*. In accordance with common practice, however, I_j will be used for the intensity of the jth reflection, the context making it clear whether I is a number of counts or a counting rate.

7.5.4. Fixed-count timing

The probability of a time t being required to accumulate N counts when the true counting rate is ν is given by a Γ distribution (Abramowitz & Stegun, 1964, p. 255):

$$p(t)\,\mathrm{d}t = [(N-1)!]^{-1}(\nu t)^{N-1}\exp(-\nu t)\,\mathrm{d}(\nu t). \tag{7.5.4.1}$$

The ratio N/t is a slightly biased estimate of the counting rate ν; the unbiased estimate is $(N-1)/t$. The variance of this estimate is $\nu^2/(N-2)$, or, nearly enough for most purposes, $(N-1)^2/(N-2)t^2$. The differences introduced by the corrections -1 and -2 are generally negligible, but would not be so for counts as low as those proposed by Killean (1967). If such corrections are important, it should be noticed that there is an ambiguity concerning N, depending on how the timing is triggered. It may be triggered by a count that is counted, or by a count that is not counted, or may simply be begun, independently of the incidence of a count. Equation (7.5.4.1) assumes the first of these.

Equation (7.5.4.1) may be inverted to give the probability distribution of the observed counting rate ν_o instead of the probability distribution of the time t:

$$p(\nu_o)\,\mathrm{d}\nu_o = [(N-1)!]^{-1}[\nu(N-1)/\nu_o]^{N-1} \times \exp\{-(N-1)\nu/\nu_o\}\,\mathrm{d}[\nu_o/(N-1)\nu]. \tag{7.5.4.2}$$

There does not seem to be any special name for the distribution (7.5.4.2). Only its first $(N-1)$ moments exist, and the integral expressing the probability distribution of the difference of the reflection and the background rates is intractable (Wilson, 1980).

7.5.5. Complicating phenomena

7.5.5.1. *Dead time*

After a count is recorded, the detector and the counting circuits are 'dead' for a short interval, and any ionizing event occurring during that interval is not detected. This is important if the dead time is not negligible in comparison with the reciprocal of the counting rate, and corrections have to be made; these are large for Geiger counters, and may sometimes be necessary for counters of other types. The need for the correction can be eliminated by suitable monitoring (Eastabrook & Hughes, 1953); other advantages of monitoring are described in Chapter 2.3.

7.5.5.2. *Voltage fluctuations*

Mains-voltage fluctuations, unless compensated, and unsmoothed high-tension supplies may affect the sensitivity of detectors and counting circuits, and in any case cause the probability distribution of the arrival of counts to be non-Poissonian. Backlash in the diffractometer drives may be even more important in altering the observed counting rates. As de Boer (1982) says, the ideal distributions represent a Utopia that experimenters can approach but never reach. He observed erratic fluctuations in counting rates, up to ten times as big as the expected statistical fluctuations. When care is taken, the instabilities observed in practice are much less than those of the extreme cases described by de Boer. Stabilizing an X-ray source and testing its stability are discussed in Subsection 2.3.5.1.

7.5.6. Treatment of measured-as-negative (and other weak) intensities

It has been customary in crystallographic computations, but without theoretical justification, to omit all reflections with intensities less than two or three times their standard uncertainties. Hirshfeld & Rabinovich (1973) asserted that the failure to use all reflections, even those for which the subtraction of background has resulted in a negative net intensity, at their measured values will lead to a bias in the parameters resulting from a least-squares refinement. This is, however, inconsistent with the Gauss–Markov theorem (see Section 8.1.2), which shows that least-squares estimates are unbiased, independent of the weights used, if the observations are unbiased estimates of quantities predicted by a model. Giving some observations zero weight therefore cannot introduce bias. Provided the set of included observations is sufficient to give a nonsingular normal equations matrix, parameter estimates will be unbiased, but inclusion of as many well determined observations as possible will yield the most precise estimates. Requiring that the net intensity be greater than 2σ assures that the value of $|F|$ will be well determined. Furthermore, Prince & Nicholson (1985) showed that, if the net intensity, I, or $|F|^2$ is used as the observed quantity, weak reflections have very little leverage (see Section 8.4.4), and therefore omitting them cannot have a significant effect on the precision of parameter estimates.

The use of negative values of I or $|F|^2$ is also inconsistent with Bayes's theorem, which implies that a negative value cannot be an unbiased estimate of an inherently non-negative quantity. There are statistical methods for estimating the positive value of $|F|$ that led to a negative value of I. The best known approach is the Bayesian method of French & Wilson (1978), who observe that ''Instead of thanking the data for the information that certain structure factor moduli are small, we accuse them of assuming 'impossible' negative values. What we should do is combine our knowledge of the non-negativity of the true intensities with the information concerning their magnitudes contained in the data.''

7.5.7. Optimization of counting times

There have been many papers on optimizing counting times for achieving different purposes, and all optimization procedures require some knowledge of the distributions of counts or counting rates; often only the mean and variance of the distribution are required. It is also necessary to know the functional relationship between the quantity of interest and the counts (counting rates, intensities) entering into its measurement. Typically, the object is to minimize the variance of some

references