7.5. STATISTICAL FLUCTUATIONS

$$\tau = T/t, \qquad (7.5.3.7)$$

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

$$\sigma^2(\tau) = T/t^2 = \tau/t. \qquad (7.5.3.8)$$

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

$$\rho = T/t - B/b, \qquad (7.5.3.9)$$

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

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):

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

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:

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

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