Appendix C: Probability and statistics

We briefly summarize here some important results from the theory of probability and statistics. The reader is referred to the references or other texts for proofs and further details [1, 2].

Consider the measurement of some quantity X. In general, measurements of X will give different results, which we denote x. The frequency with which any result for X is obtained is given by a frequency function $f(x)$. The exact form of $f(x)$ depends on the particular process under investigation. Since the quantity X must have some value, the frequency function must have the normalization

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{C.1}$$

The function $f(x)$ is also referred to as the probability distribution function.

The expectation value for any function $g(x)$ is

$$\langle g(x) \rangle = \int_{-\infty}^{\infty} g(x) f(x)\, dx \tag{C.2}$$

which is just the sum of the various possible values of $g(x)$ weighted by the probability of having that value of x. Two expectation values are particularly important for specifying the characteristics of a distribution. The mean value of x is the expectation value of x itself, or $\langle x \rangle$. This quantity is approximated by the sample mean $\bar{x}$

$$\langle x \rangle \simeq \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{C.3}$$

where n is the number of measurements. This is, of course, a measure of the central tendency of X. The second important expectation value is the

https://doi.org/10.1017/9781009290098.019 Published online by Cambridge University Press

variance

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle$$
$$= \langle x^2 \rangle - \langle x \rangle^2 \qquad \text{(C.4)}$$

which is a measure of the spread in the measurements. The square root of the variance is called the standard deviation σ.

Three frequency functions are particularly important for the matters discussed in this book. The binomial frequency function is applicable when there are only two possible outcomes of a given measurement. For example, let A denote that some event has occurred. Suppose the measurement is repeated n times and the result A is obtained x times. The probability of this occurring is

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad \text{(C.5)}$$

where p is the probability that the event A will occur. The mean and standard deviation are given by

$$\langle x \rangle = np \qquad \text{(C.6)}$$

$$\sigma = \sqrt{np(1-p)} \qquad \text{(C.7)}$$

In the limit that the number of measurements n is large and the mean is small, the binomial distribution approaches the Poisson distribution, where

$$f(x) = \frac{y^x e^{-y}}{x!} \qquad \text{(C.8)}$$

and

$$\langle x \rangle = y \qquad \text{(C.9)}$$

$$\sigma = \sqrt{y} \qquad \text{(C.10)}$$

This distribution is frequently used for the analysis of radioactive decays, since the number of potential decaying nuclei is very large, yet the total number decaying in any short time interval is small.

The normal, or Gaussian, distribution is the limit of the binomial distribution when the number of measurements is large and the probability of the event is not too small. Many types of analog measurements exhibit a Gaussian distribution around the mean value, particularly if the measurement process is subject to random errors. The frequency is given by

$$f(x) = \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right] \qquad \text{(C.11)}$$

https://doi.org/10.1017/9781009290098.019 Published online by Cambridge University Press

and

$$\langle x \rangle = a \tag{C.12}$$

$$\sigma = \mathrm{b} \tag{C.13}$$

The full width at half maximum (FWHM) of a Gaussian distribution is related to σ by

$$\mathrm{FWHM} = 2.354\sigma \tag{C.14}$$

References

[1] P. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* New York: McGraw-Hill, 1969.
[2] A. Melissinos, *Experiments in Modern Physics,* New York: Academic, 1966, Chap. 10.