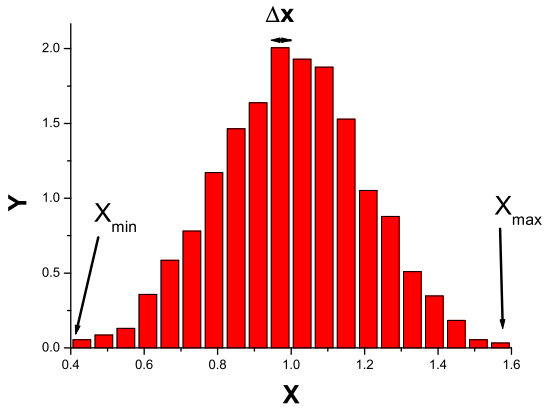

FIG. 1: Example histogram

Sometimes (as in the random number experiment) it is necessary to produce histograms from statistical data. Fig 1 shows an example histogram. The question is; how to determine the number in each interval (bin)and to turn this into code?

The number of bins used is a trade-off of resolution vs error. As the bin size is decreased the resolution of the distribution improves, but at the same time the number in each bin decreases and as a result the statistical error increases. In order to assign each random number to a bin we use the Fortran function int(X) which returns the closest integer to X rounded down. The bins are assigned to cover the expected range of X. In this case, we assume that the numbers are in the in the range $x_{min} = x_m - 3\sigma < x < x_m + 3\sigma = x_{max}$ (these values are indicated in Fig. 1). Then, if we chose n_{bin} bins, the width of each bin is $\Delta x = (x_{max} - x_{min})/n_{bin}$. Each bin is assigned an integer i in the range $(1, n_{bin})$. After each random number is generated it is given an integer value depending on the bin in which the random number lies. This is done using the fortran command *int* as follows $int((x - x_{min})/dx) + 1$, which delivers the integer label of the bin in which the random number (x) lies. The number in each bin is then stored in an array $n(i)$, with index i running from 1 to n_{bin}. This is done for n_{tot} random numbers. Note that for this exercise any number x lying outside the range can be ignored since very few will lie outside 3σ of the mean. Finally, the result of the calculation is an array containing the number of random numbers lying in each bin, whereas the Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x - x_{mean})^2/2\sigma^2) \tag{1}$$

is a normalised probability distribution, by which we mean that the probability integrated over the distribution is equal to unity, ie $\int f(x)dx = 1$. In order to compare the distribution calculated from the generated random numbers with the Gaussian distribution we calculate the probability as follows. Firstly, the number of random numbers in the range $x \to x + dx$ is

$$n = n_{tot} f(x)dx. \tag{2}$$

In terms of the discrete computer representation, we associate n with $n(i)$, dx with Δx and x with x_i where $x_i = x_{min} + (i - 0.5) * \Delta x$ is the centre of bin i. As a result,

$$f(x_i) = n(i)/(n_{tot}\Delta x) \tag{3}$$

$f(x_i)$ calculated using the random number sequence can be compared graphically with $f(x)$ calculated using equ. 1. Using $n_{bin} = 20$ Do the calculation with values of n_{tot} between 100 and 10^6; you will see the agreement with the Gaussian distribution improve with increasing n_{tot}.