

Statistical Analysis of Random Uncertainties

We know from experience that the best way to evaluate the reliability of a measurement is to repeat it several times. If the errors associated with the measurement of this quantity are random, then we can apply statistics to the multiple measurements to evaluate the uncertainty in measuring this quantity. If, however, the errors are systematic, we cannot use statistics to analyze the uncertainty.

We revisit the concept of random error with another example. Suppose we go back to the stopwatch example described earlier. We decided that there is some uncertainty associated with our ability to know when to stop the stopwatch. That uncertainty is random: we are just as likely to stop the stopwatch a little too early as we are to stop it a little too late. Over repeated measurements, this will be reflected as random error that follows a normal, or Gaussian, distribution.

Systematic errors, as previously discussed, are errors caused by instruments or some other type of repeated bias. Suppose we have designed an experiment with five trials in which we measure mass of a substance with a scale. We zero the scale before each measurement, but, unbeknownst to us, the scale doesn't zero properly. For each trial, then, the mass measured is off by the same amount, producing a systematic error. Statistics cannot detect systematic errors and should not be used in scenarios where such errors are suspected. Instead, an experimentalist should attempt to reduce systematic errors by using calibration standards and other verification techniques.

Mean and Standard Deviation

Once we are confident that systematic errors have been sufficiently minimized and that the remaining error is random, we can use statistics to describe the data set. If we take N measurements of a certain quantity x , the best estimate of the actual quantity is the mean, \bar{x} , of the measurements:

$$x_{best} = \bar{x} = \frac{\sum_{i=1}^N x_i}{N}. \quad (27)$$

We can also estimate the error, δx . To do this, we look at how far away the individual measured values are from the mean. The value $x_i - \bar{x}$ is the deviation of a particular trial from the mean (also called a residual). We can't sum the residuals, because the construction of the definition of \bar{x} causes the residuals to always sum to zero. The solution, then, is to square the residuals, sum them, and then take the square root of that sum, resulting in a value referred to as the standard deviation, σ :

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (28)$$

The standard deviation characterizes the average uncertainty of measurements x_1, \dots, x_N . It is also sometimes known as the root-mean-squared (RMS) deviation. Additionally, sometimes an alternative definition is used for standard deviation, whereby the value of N in (28) is replaced by $(N - 1)$. This second definition produces a more conservative result of σ_x , especially when the value of N is low. Much of the time, though, there is little difference between the two forms. If we are unsure of which method to use, the safest course of action is to use the form with $(N - 1)$.

Recall that the statistics of mean and standard deviation apply only to errors that follow the normal, Gaussian, or “bell-curve” distribution. By this, we mean that if you take enough measurements of a quantity x using the same method, and if the uncertainties associated with x are small and random, then all of the measurements of x will be distributed around the true value of x , x_{true} , following a normal distribution. Although we provide no justification of this claim, it can be shown that for a normal distribution, 68% of the results will fall within a distance σ_x on either side of x_{true} , as shown in Fig. 1. Therefore, if we take a single measurement, we have a 68% probability that we are within σ_x of the actual measurement. Thus it makes sense to say that for a single measurement:

$$\delta x = \sigma_x . \tag{29}$$

We are 68% sure that a single measurement is within one standard deviation of the actual value.

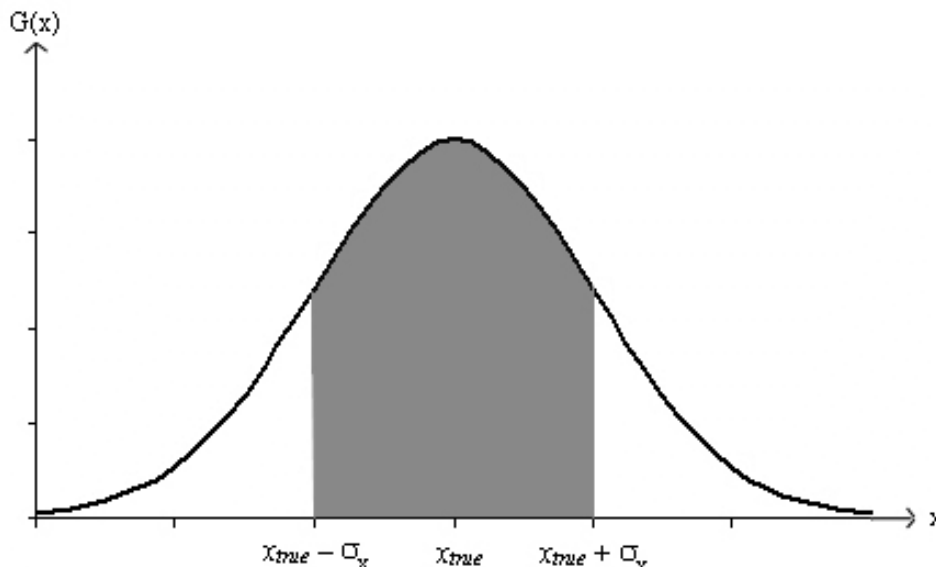


Figure. 1. A normal, or Gaussian, probability distribution. The peak of the distribution corresponds to the real value of x , x_{true} . The gray shaded area represents the area under the curve within one standard deviation of the true value. When we make a single measurement of x , there is a 68% probability that the value we measured falls in this area. Thus the stated uncertainty in a single measurement is σ_x .

Now that we know the uncertainty in one measurement, we need to know the uncertainty in the mean, \bar{x} . We expect that because the mean takes into account multiple measurements, its uncertainty will be less than that for a single measurement shown in (29). Indeed this is so. The standard deviation of the mean, or standard error, $\sigma_{\bar{x}}$ is defined as:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}, \quad (30)$$

where N is the number of measurements of the quantity x . The standard deviation of the mean is related to the uncertainty in one measurement, but is reduced because multiple measurements have been taken. It is easy to see that the more trials performed in an experiment, the smaller the uncertainty will be. Then for a series of measurements of one quantity x with independent and random errors, the best estimate and uncertainty can be expressed as:

$$(\text{value of } x) = x_{best} \pm \delta x = \bar{x} \pm \sigma_{\bar{x}}. \quad (31)$$

Adapted from chapters in “An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements” by John R. Taylor. 2nd Ed. (1997).