PC1134 Lecture 7

Topic

Error Analysis (Part I)

Objectives

Know the types of error in experimental measurements and their cause. **Understand** the fundamental theory of error analysis. **Learn** how to handle random error.

Relevance

In physical sciences, most of the experiments involve measurements of physical quantities. It is essential for one to choose the appropriate equipments and obtained a satisfactory measurement in any given experiment. Lectures 7-9 are devoted to error analysis.

Errors

Any physical quantity to be measured should in principle have a real value. However, the measured value can only be an approximation to the real value, due to the accuracy of equipment used, experimental conditions, and the skill of the observer.

The difference between the measured value and the real value is called an **error**, or **absolute error**.

In both physical measurement and data processing, effort should be made to reduce error, so that the measured quantity is as close to the real value as possible.

Experimental errors can be divided into two types, **systematic error** and **random error**, according to their characters.

Systematic Error

Systematic errors are constant errors that cannot be eliminated by simply repeating the measurement and averaging the results. They arise when something affects all the measurements of a series in an equal or a consistent way. For example,

- Incorrect calibration of an instrument,
- Construction faults in the apparatus,
- Non-constancy of experimental conditions,
- Bias of the observer,
- Incorrect use of the formulae,
- Incorrect procedure of measurement,

Systematic errors should be corrected. However, it is not easy to discover and estimate systematic errors. Most of the time, it depends on the experience and judgement of the experimentalist.

Random Error

Due to various reasons, repeated measurement of the same quantity can give different answers. This can be caused by small fluctuations in experimental conditions, reactions of the observer, etc. The measurement uncertainty or error caused by these kind of fluctuations are called **random errors**.

Random errors can be eliminated by statistical analysis. **Error analysis** deals mainly with this type of errors.

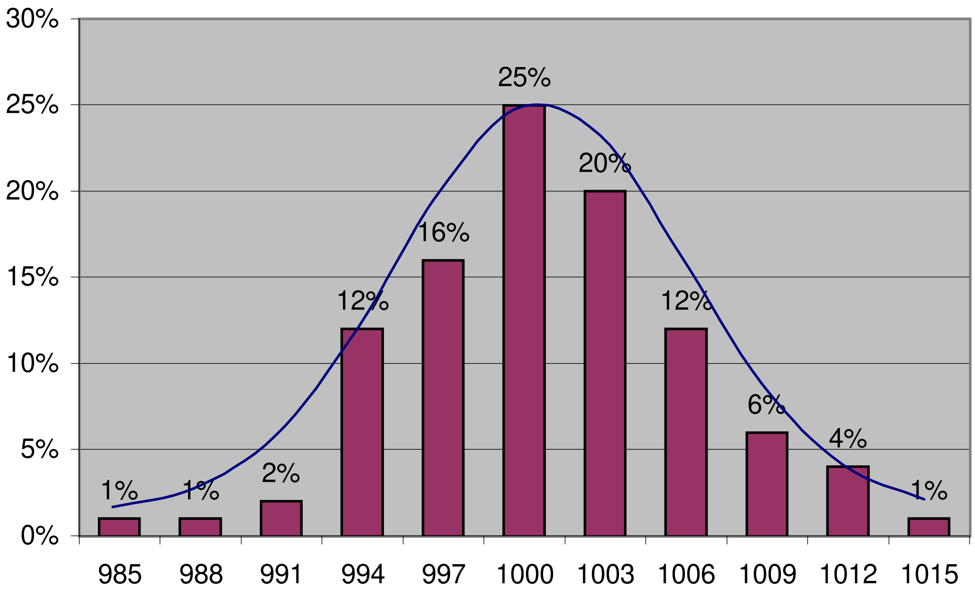
Statistics of Random Error

Random error can be reduced by repeating the measurement and applying statistics method to process the results. Assume that the following 100 values were obtained for a certain measurement.

1005.3	990.0	1001.3	1011.8	1002.5
1015.3	998.0	994.3	997.8	1003.0
996.0	1008.8	994.5	985.5	1001.3
1006.5	1001.8	1008.5	994.5	1000.0
998.5	1003.8	1001.5	1006.5	996.8
997.8	999.3	1000.0	998.8	1001.5
1003.5	1004.3	992.8	1000.3	1004.8
993.3	996.8	1002.3	1003.5	1000.0
1001.0	1006.0	1007.5	1008.8	990.0
997.0	1000.8	1004.3	1010.0	1002.8
1001.5	1000.8	998.3	1005.0	1001.5
1012.3	997.3	1000.3	1000.0	994.5
1006.5	995.0	1000.8	997.8	1009.3
1005.5	999.5	1005.8	1002.0	1001.0
994.5	1000.5	999.5	1003.8	1012.3
1001.0	1000.5	993.8	997.3	1004.5
998.8	999.3	994.8	1005.0	987.5
997.3	1002.8	993.5	998.0	993.3
998.3	997.0	999.0	1011.0	1007.3
998.3	1003.0	1003.3	1000.3	1002.3

Histogram

Range	Occurrence	Probability
$\leq$ 986.5	1	1%
986.6 – 989.5	1	1%
989.6 – 992.5	2	2%
992.6 – 995.5	12	12%
995.5 – 998.5	16	16%
998.6 – 1001.5	25	25%
1001.6 – 1004.5	20	20%
1004.6 – 1007.5	12	12%
1007.6 – 1010.5	6	6%
1010.5 – 1013.5	4	4%
$\geq$ 1013.6	1	1%



Distribution

In general, in a total of N measurements, if there are n_i values in the interval $[x_i - \Delta x/2, x_i + \Delta x/2]$, we say that the probability of finding the values at x_i is

$$p(x_i)\Delta x_i = \frac{n_i}{N}$$

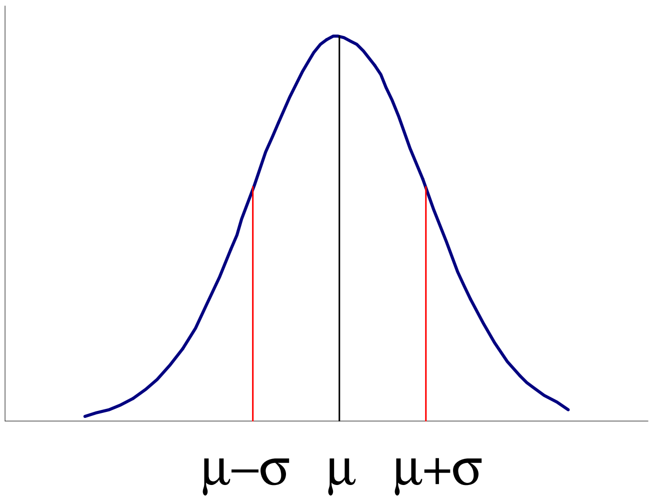
As the number of measurements go to infinity and the interval Δx goes to zero, we then obtain a smooth probability distribution function.

For most of measurements, the distribution is given by the **Gaussian distribution** or **normal distribution**.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}$$

where x is a measured value, μ is the peak value of the distribution, σ is a real constant which is called the **standard deviation**.

Normal Distribution



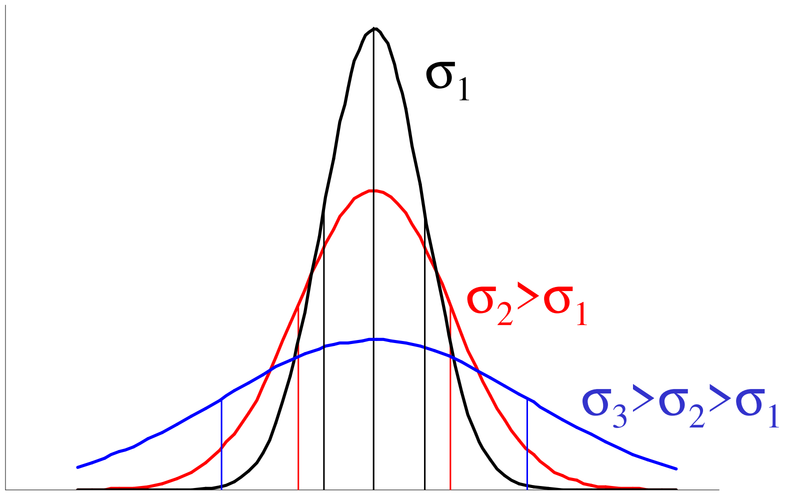
If μ and σ are known, then the probability of finding x in the interval $[a, b]$ is given by

$$P(a < x < b) = \int_a^b p(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-(x-\mu)^2/2\sigma^2} dx$$

In particular,

$$\begin{aligned} P(\mu - \sigma < x < \mu + \sigma) &= 68.3\% \\ P(\mu - 2\sigma < x < \mu + 2\sigma) &= 95.4\% \\ P(\mu - 3\sigma < x < \mu + 3\sigma) &= 99.7\% \\ P(\mu - 1.96\sigma < x < \mu + 1.96\sigma) &= 95\% \end{aligned}$$

Standard Deviation



Small σ, $\Rightarrow$ narrow distribution of data and relatively accurate measurement.

Large σ, $\Rightarrow$ large spread of data points and large uncertainty in measurement.

Average

Assume that any systematic error in a measurement of μ has been removed. The random error of a measured value x_i is

$$\delta_i = x_i - \mu$$

Sum over all measured values and divide both sides of the equation by N, we get

$$\frac{1}{N}\sum_{i=1}^{N}\delta_i = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu) = \frac{1}{N}\sum_{i=1}^{N}x_i - \mu$$

Since the random errors tend to cancel each other, when $N \to \infty$, $\sum \delta_i \to 0$. Therefore,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i \to \mu$$

The average of all measured values, $\bar{x}$, is a good estimate to the real value, μ.

Standard Deviation

The uncertainty of a single measurement is characterized by the standard deviation.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Since the real value μ is unknown, we use $x_i - \bar{x}$ instead of $x_i - \mu$.

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

where $x_i - \bar{x}$ is called a **residual error**.

It can be shown that when N is sufficiently large, the two definitions produce the same result. It can also be shown that the σ given above is the same as that appears in the normal distribution and the average $\bar{x}$ is the same as the x value corresponding to the peak of the normal distribution.