EC-1(Measurement and Instrumentation)Unit 1 Notes (II)

Amarendra Narayan

Treatment of Experimental errors

Uncertainties of measurements: Experiments often generate multiple measurements of the same thing, and these measurements are subject to error. Statistical analysis can be used to summarize those observations by estimating the average, which provides an estimate of the true mean. Statistical calculation for summarizing the observations also give the variance in the experimental data, which describes the uncertainty in the measured variables. Sometimes we have made measurements of one quantity and we want to use those measurements to infer values of a derived quantity. Statistical analysis can be used to understand the propagation of the measurement error through a mathematical model and to estimate the error in a derived quantity.

Types of experimental errors: Experimental errors can be of mainly two types: Systematic errors and Random errors.

Systematic errors occur in the output of insts possibly due to factors inherent in the manufacture of the instrument such as those arising due to tolerances in the components of the instrument. They can also arise due to wear and tear of the instrument components over a period of time. Systematic errors could also arise due to environmental disturbances or through the disturbance in the measurand system introduced by the act of measurement.

Random errors occur in measurements because of unpredictable variations in the measurement system. They are usually observed as small perturbatiotrumenns of the measurement on either side of the correct value. Positive errors and negative errors occur in approximately equal numbers if a series of measurements of the same quantity is done. Therefore, random errors can largely be eliminated by calculating the average of a number of repeated measurements, provided that the measured quantity remains constant during the process of taking the repeated measurements.

Absolute and relative error: There are two common ways in which errors of measurement are expressed. When the error is expressed as a value, for example 3.5 cm ± 0.35 cm then the value of error is ± 0.35 cm. Error expressed this way (as a value) is called absolute error. This error can also be expressed in another way – as a fraction of the expressed mean value. 0.35 cm is 10% of 3.5 cm. Thus we can express the same quantity together with error as 3.5 cm ± 10%. The error here is 0.35 cm which is 10% of the mean value of 3.5 cm. Error expressed this way is called relatie error.

Propagation of errors: When some measured value is used in calculation of some other quantity, the errors in curred during original measurement get transferred and appear as error in derived quantity. This is known as propagation of errors. A simplistic approach is often used to obtain Limiting error (maximum error) in derived quantity using the following guidelines.

Error in sum or difference: Suppose we are interested in finding out error in sum of two values X_1 and X_2 having errors respectively $\pm\Delta X_1$ and $\pm\Delta X_2$. The sum of their mean values is $X_1 + X_2$ and their difference is $X_1 - X_2$. Since we do not know whether the error is positive or negative, so whether it is the sum or the difference, the error in the worst case will be $\pm(\Delta X_1 + \Delta X_2)$. So in order to obtain limiting error in sum or difference, *we add the absolute errors*.

Error in product or quotient: Suppose we are interested in finding out error in product of two values X_1 and X_2 having errors respectively $\pm\Delta X_1$ and $\pm\Delta X_2$. The product of their mean values is $X_1 \times X_2$ and their quotient is X_1 / X_2 . It can be shown using simple calculus that for product and quotient, the relevant limiting error is not the absolute error but the relative error. Also, we do not know whether the error is positive or negative. So whether it is the product or the quotient, the error in the worst case will be $\pm\left(\frac{\Delta X}{X}+\frac{\Delta y}{Y}\right)$. Thus in order to obtain limiting error in the product or the quotient, *we add the relative errors*.

Working with limiting errors, we generally obtain worst case values that are unrealistically large. It is more pragmatic to assume that the errors are random and helse follow a gaussian distribution. The following section describes the treatment of errors for normally distributed data.

Combination of errors in Normally distributed data

When we make a single measurement of a quantity then in a majority of situations it is not possible to ascertain the correctness or reliability of that measurement. So wherever possible we repeat the measurements. In repeated measurements the obtained values are distributed about some mean value and we can take the mean value as a better estimate of the actual value as against tha value obtained from any single measurement. MeasurementX values vary when we make repeated measurements. Variation in measurement values can arise due to several possible reasons. For example, there can be difficulties in taking reading, or making judgment of the most appropriate value, or there might be a change in experimental conditions from one reading to another that changes the measured value, or else the parameter to be measured itself might have inherent random fluctuations. For these reasons we take a set of readings and deduce the value from it using statistical analysis. It is a good practice in experimental work to repeat the measurement several times so as to ensure better accuracy in measurement.

The result of a set of repeated measurements can be summarised with help of certain parameters which are given below:

Suppose we have a set of N values [x_1, x_2, ...x_i, ... x_n]

The average of these values is expressed in terms of **mean.** The mean of these values, $\bar{x}$ is,

$$\bar{x}=\frac{1}{N}\sum(x_1+x_2+...+x_i+...+x_n)=\frac{1}{N}\sum_{i=1}^{n}x_i$$

The variation in the set of values is usually expressed as **sample variance** s^2 which is given by

$$s^2=\frac{1}{(N-1)}\sum(x_i-\bar{x})^2$$

If the number of readings N is very large then (N - 1) can be approximated by N in the above expression for variance and we have

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

where σ is called the **standard** deviation of the population.

The variance is a parameter that describes the variations to be expected in individual measurement values. But how certain can we be about the mean that we have obtained from a finite set of readings that we have taken during the experiment? The **standard deviation of mean** is a parameter that describes the variability of the mean – What is the typical variation in the computed mean that we should expect if we repeat the whole experiment. The variability of the mean is given by s_m^2, the variance of the mean.

$$s_m^2=\frac{1}{N(N-1)}\sum(x_i-\bar{x})^2=\frac{s^2}{N}$$

Distribution Function: A Distribution function helps in summarizing a set of data. Its plot is a useful tool for visualizing the data set. In order to create a distribution function from a data set, we first determine the range of the data which extends from the minimum value x_{min} to the maximum value $x_{max.}$ This range is divided into a number of equally spaced intervals. Each interval will thus cover a smaller range (subrange). If there are N intervals then each interval has a subrange that is N times smaller, i.e.

$$\frac{x_{max}-x_{min}}{N}$$

Next we count the number of data values that lie within each subrange. We will obtain a frequency distribution. This frequency distribution for a finite set of data is called a **sample distribution.** It can be plotted in the form of a **Histogram**.

Example:

The Parent Distribution: If the number of data values is very large, we can construct a larger number of subranges, and still assure that there is sufficient number of data values in each subrange. The subranges are now smaller and each subrange also has sufficient number of data values. If we assume that our histogram is derived from some parent distribution, then in the limiting case as $N \rightarrow \infty$ the histogram will tend to a smooth curve which will actually be the **parent distribution**. Where the number of data values is finite, the values in the subintervals are scattered about the line defining the parent distribution. The parent distribution is a continuous distribution. So the probability of occurrence in any range $[x_1 \leftrightarrow x_2]$ is given by integrating a probability density function over x in that range.

Probability Distribution: A Probability distribution is a distribution of probabilities of the value lying in different subranges. If in the set of data values of variable *x,* a certain range is chosen and the number of data values in that range is divided by the total number of data values obtained during the experiment, then it expresses the probability of occurrence of a value within that range, where $n_i = n(x_i)$. is the number of values in the range x_i to $x_i + \Delta x$. The limiting case of this probability distribution as $i \rightarrow \infty$ is $P(x, n)$ which represents the probabilities for the parent distribution. The mean in the parent distribution is called μ and the variance σ^2. The standard deviation is σ.

The distribution function as described above has been obtained from a set of data. Note that a distribution function for a set of measurements does not have any pre-defined shape and only depends on the nature of the experimental data.

We can also obtain a distribution function based on some theoretical model which is based on some particular assumptions. Such theoretical models are handy and are often used for statistical predictions.

Three most commonly used distributions are (1) Binomial distribution, (2) Poisson distribution, and (3) Normal or Gaussian distribution.

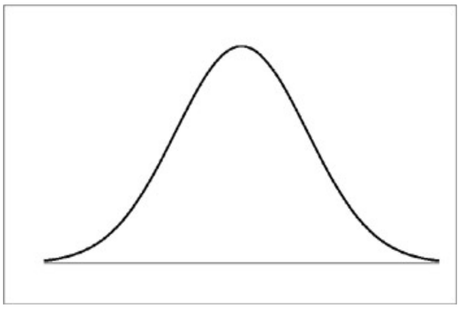
Normal Distribution: Normal Distribution is the distribution that is most commonly observed in Nature. Most of the continuous variables – the probability follows a normal distribution. Examples are Height, weight, etc. This is one of the most important concepts of Mathematics.

We can express the Normal distribution as:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

Here $\exp(x)$ stands for e^x.

2, π, e are constants, x is a variable, and μ and σ are parameters for this distribution.



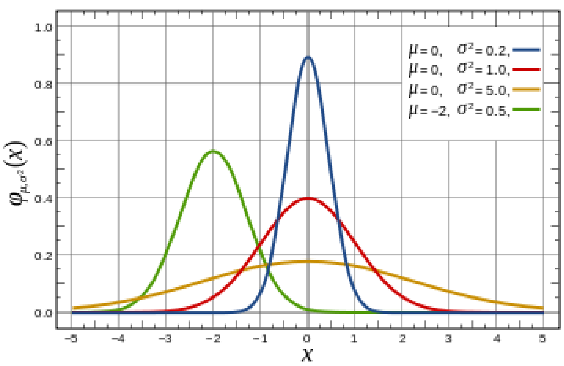
μ is the mean of the distribution, and σ^2 is called variance which is a measure of variability or spread in data. Corresponding to different values of μ and σ we have different normal distributions. The normal curve is bell shaped and is symmetric about the mean. That is, *f(x) = f(-x)* where x is the deviation from the mean.

Examples:

(1) $f(x)=\frac{1}{2\sqrt{2\pi}}\exp\left(\frac{-(x-4)^2}{8}\right)$ is a normal distribution with mean = 4 and σ = 2.

(2) $f(x)=\frac{1}{2\sqrt{\pi}}e^{\frac{-x^2}{4}}$ is a normal distribution with mean = 0 and $\sigma^2 = 2$.

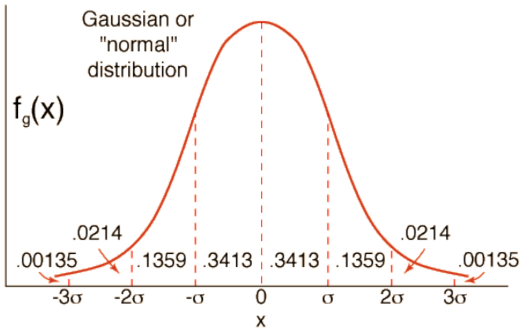
(3) $f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ is a normal distribution with mean = 0 and σ = 1.



Thus there are infinite number of possible distributions for different μ and σ, out of which the distribution with μ= 0 and σ = 1 is called Standard Normal Distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

(Whenever in Statistics we use the term standard, it generally means that the mean is 0 and variance is 1. In a standard normal curve a particular deviation from mean is usually called Z.



The curve corresponding to a normal distribution is called a normal curve. A normal curve is a symmetric bell-shaped curve that extends from -∞ to ∞. For a normal curve μ represents mean, median as well as mode. For a standard normal curve, mean = median = mode = 0, and total area under the curve = 1.

In discrete distributions, we measure probabilities of values, but in continuous distributions, we measure probabilities of intervals. The main reason why we are interested in normal distribution is that its applicability is quite wide and wherever applicable, it is a convenient tool for obtaining probability in case of continuous distributions.

For a given distribution, we define **confidence level** for a given interval as the probability that the value of the given parameter lies within the specified interval. Some important and often used values are:

Interval	Confidence level
-σ ↔ +σ	68.00%
-2σ ↔ +2σ	95.00%
-3σ ↔ +3σ	99.50%
-0.675σ ↔ + 0.675σ	50.00% (most probable value)
2.35σ	Full width at Half maxima

We notice that wherever we have a normal (gaussian) distribution the majority of data values

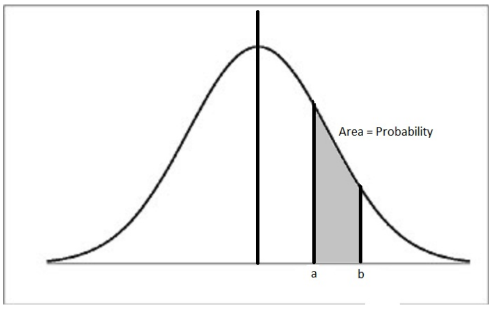
(99.5%) will lie within ±3σ of the mean. Also, half of all data will lie within ± 0.675σ of the mean and half of the data will lie outside this interval. The range from – 0.675σ to +0.675σ is called most probable value. The **Full Width at Half Maxima** (FWHM) is 2.35σ.

Finding Probability of a continuous variable having Normal distribution:

For continuous variables, probability means area under the curve.

$$P(a<x<b) = \text{Area under normal curve from } x = a \text{ to } x = b$$

The total area under a standard normal curve is 1. Since the total probability of occurrence of any event is also 1, we can equate the two and say that the a standard normal curve can be used to obtain probability of occurrence in case of events that have a normal distribution.



A normal table is a table in which area under a standard normal curve from 0 to Z is tabulated for different Z. Using a normal table we can obtain the probability of occurrence of an event within a given interval expressed in terms of Z. Such tables are available in two forms: One form gives integrated area from - ∞ to z (Generally for negative z) and the other form gives integrated area from 0 to z (Generally for positive z). Since the standard normal curve is symmetric about z, we can use either of the two forms with a small difference in calculation method.

Examples:

Realistic situations do not follow standard normal curves. How do we translate the problem for non standard normal distribution curves?

$$x = N(\mu, \sigma^2)$$

is a normal variate. We need to transform x into a standard normal variate.

$$z = \frac{x-\mu}{\sigma}$$

is a standard normal variate N(0,1) that corresponds to x. In order to find probabilities in case of non standard normal variate x, we transform it into a standard normal variate z and follow a similar procedure as before.

The Normal Gaussian Law of Errors:

The function defining a Normal frequency distribution is also featured in the **Gaussian Law of Errors.** This Law states that when we measure a certain physical quantity which is subject to accidental random errors, then the measured values are distributed about the mean value according to a normal distribution. In other words, any set of measurements of a given quantity can be considered as a sample taken from a very large population – the sumtotal of all measurements that could be done if the instrument and time permitted it, and this population is normally distributed.

The two parameters that define the distribution of the population are the population mean μ and the standard deviation in the population, σ. Now, two questions are relevant here. How can we obtain these parameters for the whole (infinite) population from a limited set of observations? If we do calculate the values of these parameters, then what will be the expected error in such an estimate? The best estimate of the value of the measured quantity is the sample mean $\bar{x}$. The accuracy of this value is the standard error of the mean s which can be obtained from the standard deviation of the sample a σ as $s = \sigma/\sqrt{n}$ where n is the number of measurements. Sometimes the probable error of the mean, $0.675\,\sigma/\sqrt{n} \simeq 2/3\,s$ is used to denote the accuracy of obtained mean.

Please note that the normal law of errors is based on certain theoretical assumptions that may or may not be true in a particular case. The assumption in this law is that the deviation of any particular measurement from the mean value is due to one or more of a very large number of possible small deviations that can occur due to numerous independent causes and that there is equal probability of a deviation being positive or negative. This law is not universally valid. But it is very useful in cases where we are confronted with random errors or errors from unknown causes. The normal law cannot be theoretically proved. But it has been generally found to be not far wrong, and also, it is easy to handle.

Reproductive property of the normal error law:

Any deviation which can be expressed as the sum of a set of deviations, each of which satisfies the normal error law, will itself satisfy the normal error law. Thus any linear combination of normal distributions is also a normal distribution. This is known as the reproductive property of normal error law.

Suppose that a deviation D is a linear function of n independent deviations.

$$D = k_1 d_1 + ... + k_n d_n \quad , \qquad k_{1..}k_n \text{ are constants.}$$

Let each of these deviations d_s satisfy the normal distribution law.

The probability of d_s lying between x and $x + dx$ is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

Since $D = k_1 d_1 + ... + k_n d_n$, the probability of D lying between x and $x + dx$ is given by

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

where $$\sigma^2 = k_1^2 \sigma^2 + ... + k_n^2 \sigma_n^2$$

Proposition: Let $X_{1,} ..., X_n$ be independent normally distributed random variables with no assumption made on the means and variances. Then any linear combination of these random variables is again normally distributed.

Propagation of Error:

Most of the measurements are indirect measurements. Indirect measurements are done where it is either not possible or not practical to measure the physical quantity directly. In such cases of measurement, we measure a set of parameters that are functionally related with the quantity, and either use a calibration or a mathematical formula expressing the relation to obtain the measured value. How does the errors in directly measured parameters affect the value of the quantity we are interested in? In other words,how does error propagate during indirect measurements? Let us see.

Consider an experiment in which we want to obtain a measurement of a quantity u which is a function of some directly measurable quantities, $x, y, ...$ $u = f(x, y, ...)$ Thus Some examples are, $Volume = length \times breadth \times height$, or $E = kT \ln(R_T / R_0)$.

In designing any experiment, we must take into consideration the possible errors as well as relative errors that we can allow for each of the parameters which are measured.

Suppose we are interested in measuring a quantity u where $x, y, ...$ are the directly measurable quantities, and

$$u = f(x, y, ...)$$

The variance of u is given by

$$\sigma^2 = \lim_{n \to \infty} \sum_{i=1}^{n} (u_i - \bar{u})^2$$

We expand $(u_i - \bar{u}) = f(x_i, y_i, \ldots)$ in a Taylor series

$$(u_i - \bar{u}) = (x_i - \bar{x}) \frac{\partial u}{\partial x} + (y_i - \bar{y}) \frac{\partial u}{\partial y} + \ldots$$

and obtain the variance

$$\sigma^2 = \lim \sum \left[(x_i - \bar{x}) \frac{\partial u}{\partial x} + (y_i - \bar{y}) \frac{\partial u}{\partial y} + \ldots \right]^2$$

When take the square the expression in the brackets, we find that there are two types of terms:

$$(x_i - \bar{x})^2 \left(\frac{\partial u}{\partial x} \right)^2 = \sigma_x^2 \left(\frac{\partial u}{\partial x} \right)^2$$

and

$$(x_i - \bar{x})(y_i - \bar{y}) \left(\frac{\partial u}{\partial x} \right)^2 = \sigma_x^2 \left(\frac{\partial u}{\partial x} \frac{\partial u}{\partial y} \right)$$

The second term is called the **covariance**. It is 0 if x and y are independent measurable parameters and are unrelated. Equating it to 0 we obtain the following error propagation formula:

$$\sigma_u^2 = \sigma_x^2 \left(\frac{\partial u}{\partial x} \right) + \sigma_y^2 \left(\frac{\partial u}{\partial y} \right) + \ldots$$

Using this formula we can obtain the variance of different combined parameters.

(1) **Variance of Sum or Difference:**

We have, $x = \bar{x} \pm \sigma_x$, $y = \bar{y} \pm \sigma_y$, $\frac{\partial u}{\partial x} = 1$, $\frac{\partial u}{\partial y} = 1$

$$\sigma_u^2 = \sigma_x^2 (1)^2 + \sigma_y^2 (1)^2$$

or

$$\sigma_u = \sqrt{\sigma_x^2 + \sigma_y^2}$$

(2) Variance of a Linear combination:

Let $u = a_1 x_1 + a_2 x_2 + \ldots$

We have, $x_1 = \bar{x}_1 \pm \sigma_{x1}$, $x_2 = \bar{x}_2 \pm \sigma_{x2}$, $\frac{\partial u}{\partial x_1} = a_1$, $\frac{\partial u}{\partial x_2} = a_2$ and so on...

$$\sigma_u^2 = \sigma_{x1}^2 (a_1)^2 + \sigma_{x2}^2 (a_2)^2 + \ldots$$

or $$\sigma_u = \sqrt{a_1^2 \sigma_{x1}^2 + a_2^2 \sigma_{a2}^2 + \ldots}$$

(3) Variance of a mean

Let $$u = \bar{x} = \frac{1}{N} \sum (x_1 + x_2 + \ldots + x_n)$$

$$\frac{\partial u}{\partial x_1} = \frac{1}{N}, \ldots \frac{\partial u}{\partial x_n} = \frac{1}{N}$$

$$\sigma_u^2 = \frac{1}{N^2} (\sigma_1^2 + \sigma_2^2 + \ldots) = \frac{1}{N^2} N \sigma^2 = \frac{\sigma^2}{N}$$

(4) Variance of product:

Let $$u = axy$$

$$\frac{\partial u}{\partial x} = ay \quad \text{or} \quad \frac{\partial u}{axy} = \frac{\partial u}{u} = \frac{\partial x}{x}$$

$$\text{and} \quad \frac{\partial u}{\partial y} = ax \quad \text{or} \quad \frac{\partial u}{axy} = \frac{\partial u}{u} = \frac{\partial y}{y}$$

$$\sigma_u^2 = \sigma_x^2 \left(\frac{\partial u}{\partial x} \right) + \sigma_y^2 \left(\frac{\partial u}{\partial y} \right) = \sigma_x^2 (ay)^2 + \sigma_y^2 (ax)^2$$

$$\text{or} \quad \frac{\sigma_u^2}{u^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2}$$

(5) Variance of Quotient:

Let $$u = a \frac{x}{y}$$

$$\frac{\partial u}{\partial x} = \frac{a}{y} \quad \text{or} \quad \left(\frac{y}{ax} \right) \partial u = \frac{\partial u}{u} = \left(\frac{y}{ax} \right) \partial x \left(\frac{a}{y} \right) = \frac{\partial x}{x}$$

$$\text{and} \quad \frac{\partial u}{\partial y} = \frac{-a}{x^2} \quad \text{or} \quad \left(\frac{y}{ax} \right) \partial u = \frac{\partial u}{u} = \left(\frac{y}{ax} \right) \left(\frac{-a}{y^2} \right) \frac{\partial y}{y}$$

$$\sigma_u^2 = \sigma_x^2 \left(\frac{\partial u}{\partial x} \right) + \sigma_y^2 \left(\frac{\partial u}{\partial y} \right) = \sigma_x^2 (ay)^2 + \sigma_y^2 (ax)^2$$

$$\text{or} \quad \frac{\sigma_u^2}{u^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2}$$

or $$\frac{\sigma_u}{u}=\sqrt{\frac{\sigma_x^2}{x^2}+\frac{\sigma_y^2}{y^2}}$$

Other Formulae that can be similarly obtained:

u	σ_u or $\frac{\sigma_u}{u}$	Comment
$x \pm y$	$\sigma_u=\sqrt{\sigma_x^2+\sigma_y^2}$	Standard deviations add in quadrature
$ax \pm by$	$\sigma_u=\sqrt{a^2\sigma_x^2+b^2\sigma_y^2}$	
$\pm axy$	$\frac{\sigma_u}{u}=\sqrt{\frac{\sigma_x^2}{x^2}+\frac{\sigma_y^2}{y^2}}$	Relative Standard deviations add in quadrature
$\pm a\frac{x}{y}$	$\frac{\sigma_u}{u}=\sqrt{\frac{\sigma_x^2}{x^2}+\frac{\sigma_y^2}{y^2}}$	
$x^{\pm m}$	$\sigma_u/u=\pm m\sigma_x$	
$e^{\pm ax}$	$\sigma_u/u=\pm a\sigma_x$	
$a^{\pm bx}$	$\sigma_u/u=\pm(b\ln a)\sigma_x$	
$\ln(\pm bx)$	$\sigma_u/u=b\sigma_x/x$	
$\bar{x}=\frac{1}{N}\sum x_i$	$\sigma(\bar{x})=\sigma/\sqrt{N}$	If all standard deviations are equal to σ
$\bar{x}=\frac{\sum(x_i/\sigma_i^2)}{\sum(1/\sigma_i^2)}$	$\sigma(\bar{x})=\frac{1}{\sum(1/\sigma_i^2)}$	If each reading x_i has standard deviation σ_i

Deviations from normal distribution for small samples

Normal Distribution is based on certain assumptions. The first assumption is that the data is randomly distributed. (If there are many possible unrelated reasons for variations from one observation to another and neither these reasons nor their nature is known from before, then in such cases we say that the errors are randomly distributed.) The second assumption is that the number of observed data instances is very large (It tends to infinity).

In any real life situation neither of these assumptions would be valid. The assumptions regarding random distribution are a proclamation of our ignorance regarding the behavior of the system under study – we are not understanding the behavior and yet we want to make predictions about the behavior. There is actually no valid reason for doing so except that such a prediction works if we assume a normal distribution. All the different frequency distributions tend to a normal distribution when the number of data points gets very large. If an observed distribution does not do so, then we can usually find out the reasons for such deviations from a normal distribution and say that the deviations are due to systematic errors. Then we proceed towards finding these systematic errors. Sometimes we are successful, sometimes we are not. But in cases where the distribution is close to normal and there are a sufficient number of data, then a random distribution with a gaussian shape is a good assumption.

The second situation is when the number of observed data is small (less than 30). Then there is a deviation with respect to the normal curve. The appropriate distribution for small sample sizes is different because in obtaining normal distribution we had made an approximation that is no more valid. We had approximated s^2 with σ^2 where $s^2=\frac{1}{(N-1)}\sum(x_i-\bar{x})^2$ and $\sigma^2=\frac{1}{N}\sum(x_i-\bar{x})^2$. For N < 30 there is a significant difference in the two curves represented by $f(x)=\frac{1}{s\sqrt{2\pi}}\exp\left(\frac{-(x-\bar{x})^2}{2s^2}\right)$ and $f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$. The latter represents the gaussian distribution. The first one is called **Student's** t **distribution.**

Distribution functions
Derivation of Distribution functions
Properties of Distribution functions