ANALYSIS OF RANDOM ERRORS IN PHYSICAL MEASUREMENTS: A PRIMER

by **Dr. Pramod Joag**

Department of Physics

University of Pune

Pune – 411007

Contents

1 Preamble **1**

2 Random Errors **2**

3 Aims and Objectives **3**

4 Random and Systematic Errors **4**

5 Histograms (Frequency distributions) **5**

6 The Limiting Distribution **6**

7 The True Value Revisted **8**

8 Mean and Standard deviation for $p(x)$ **9**

9 The Normal (Gaussian) Distribution **9**

10 Interpretation of σ as the Measure of Uncertainty **11**

11 Evaluation of the Best Estimate for x & σ **12**

12 Two Important Questions **14**

13 How to Report your Results **15**

14 The Propagation of Errors **16**

15 Standard Deviation of the Mean (SDOM) **17**

16 Confidence **19**

17 The Weighted Average **20**

18 Rejection of Data (Chauvenet's criterion) **22**

19 The Poisson process **23**

20 Some Properties of the Poisson Distribution **27**

21 Examples **29**

22 The χ^2 test for the Goodness of fit **29**

23 The Lorentzian distribution; an example **31**

24 The General Definition of χ^2 **32**

25 Degrees of Freedom and Reduced χ^2 **32**

26 The probabilities for χ^2 **34**

27 Limitations of the classical approach **35**

Appendix I **37**

Appendix II **40**

Appendix III **43**

1 Preamble

Measurement of various atttributes of a system is a fundamental process for all natural sciences, and most certainly for physics. Now, it turns out that any physical attribute (position, momentum, energy, voltage, current etc.) of any physical system cannot be measured to arbitrarily high precision, *even in principle*. Thus, if you measure a physical quantity x pertaining to a particular system, and you get a value x_0 as a result, then x_0 cannot be regarded as exactly the same as the value of x possessed by the system at the time of measurement (say x_{true}). If we have a reliable estimate of the uncertainty involved, we can say that x_{true} lies in the range,

$$x_0 + \delta x > x_{true} > x_0 - \delta x \tag{1}$$

where δx is the uncertainty or the error in the measured value.

Before analysing the reasons for these unavoidable uncertainties or errors in the measurements, let us have a word about the kind of physical systems and their physical attributes we wish to talk about. We shall adopt the following point of view. At every instant of time, the physical system possesses a unique and definite value for all the physical attributes which can be measured, independent of whether we actually perform any measurement or not. These values are also independent of our method of measurement or the kind of apparatus used. Thus, for example, at every instant of time a material particle is in a state characterised by definite values of its position, momentum, angular momentum and so on.

For a given combination of the physical quantity say x, and the system under consideration, we call the definite value of x possessed by the system at a given time to be the true value of x and denote it by x_{true}. Furthermore, these true values of various physical quantities, pertaining to the classical system, are smooth functions of the given external conditions and time. Therefore, if the state of the system at any given instant is known (by measuring the relevant physical quantities) the complete temporal development of the system can be exactly predicted (both for the future and the past instants). The above point of view and the corresponding measurements are called classical or deterministic. In this note we will be concerned with deterministic measurements, unless otherwise stated. We emphasise that not all physical systems yield to the classical approach. We will touch upon this point at the end of this note.
Here are a few examples of the physical quantities we are interested in:

1. Universal constants like velocity of light in vacuum, charge of an electron, acceleration due to gravity (we assume that this is constant).

2. Various electrical quantities like voltages, steady currents, capacitance, inductance and so on.

3. Equilibrium thermodyanmic quantities like pressure, temperature, volume, entropy, heat capacities etc.

4. Bulk properties of materials e.g. the resisitivity of highly pure single crystal of germanium, in a given crystallographic direction, at temperature = 77 °K.

We shall give an example of the process which is not deterministic. Suppose a particle, large enough to be seen by the naked eye, is introduced in a liquid at $t = 0$. It will suffer collisions with liquid molecules which are enormous in number, and consequently will trace a path in the liquid as time passes. Note that it is impossible to account for the elementary collision events in detail. You can't do anything more than look at the motion of the particle itself. In such circumstances, the path traced by the particle is completely random i.e. by knowing the position of the particle at time t, you cannot predict exactly its position at a subsequent instant. All that you can do is seek the probabilities for the various values of the particle's position at any given time. Obviously, the position of the particle does not have a definite value at any given time. The motion described above is called the Brownian motion.

2 Random Errors

The fundamental reason for this uncertainty or error mentioned in the first paragraph is that any measuring apparatus and associated equipment is made of some material and that material is at a finite temperature. The atoms and electrons that make up this material jostle around in a random manner, and because of this they give rise to a signal that is essentially indistinguishable from the signal that you are trying to measure.Thus, the observed value is essentially uncertain, or erroneous, by the value of the above mentioned 'noise signal' (The term 'noise' means any random i.e. unpredictable process or motion), which gets added up to the signal due to the system under observation. Since the value of the noise signal fluctuates randomly in time, so also does the corresponding error in the observed value. Therefore, it is called a 'random error'. Ofcourse, if your measured value has to have any meaning, the magnitude of the physical quantity being measured must be much greater than the magnitude of the noise signal i.e. the signal to noise ratio (S/N) must be as high as possible.

It is to be emphasised that the random errors, introduced by the random fluctuations of atoms comprising the apparatus, are far smaller than many other sources of random and systematic errors (see article 4) which are more important in practice. These random fluctuations set up the lowest limit for the detectability of a signal or the highest limit on the accuracy with which a measurement can be made.

In practice, the random errors can show up in the following (and the only) way. If you repeat your measurement under identical conditions at various times, the random error

in the measurement will take a value at random every time you repeat your measurement. Therefore, the observed value, which contains the random error, will also fluctuate randomly i.e. you cannot predict the observed value at the n^{th} measurement by knowing any or all the values at the previous measurements. However, the observed values will be found to lie in a finite range.

3 Aims and Objectives

Before proceeding with our analysis, let us state clearly our aims and objectives. Suppose, we repeat the measurement of a physical quantity say x, under identical conditions, a large number of times so that we have a large amount of data revealing the random errors in our measurements. What do we do with it? Well, concealed in our data is the information about the true value of x, x_{true} i.e. the value of the quantity x possessed by the system under observation at the time of measurement, and also about the range of the random errors involved. Thus on the basis of the observed data, we want to find the best estimate of x_{true} (call it x_{best}) and the best estimate of the amount of uncertainty in the value of x_{best}. That is we must be able to write

$$(\text{measured value of } x) \;=\; x_{best} \pm \delta x \tag{2}$$

δx is the amount of uncertainty in the measurement of x and has the following meaning: Your apparatus cannot meaningfully differentiate between two values of x say x_1 and x_2 if $|x_1 - x_2| \leq \delta x$ *

Since the observed values fluctuate randomly the best thing we can hope to know is the probability of occurence of any particular value, relative to that of the others. In practice, if want to find these probabilities we will have to repeat the measurement, under identical conditions, infinitely many times. This is clearly an idealization. Thus we are faced with the following problem: With the finite amount of data (however large) available to us assuming that it follows a particular probability distribution, how to obtain the best estimate of the true value of the physical quantity being measured and also of the amount of this random error. Finally, we would like to test how good was this hypothesis that our data fits a given probability distribution.

*Suppose, we make two measurements of the same quantity x at two different times, on the same physical system to get values x_1 and x_2 respectively. Then if $|x_1 - x_2| \gg \delta x$ it is overwhelmingly probable that x_1 & x_2 are the values possessed by the system at the two times of measurement rather than x_1 & x_2 being due to random errors in measurement. If $|x_1 - x_2| \leq \delta x$ we cannot say anything about x_1 & x_2 being two values possessed by the system under observation, we say that these values cannot be resolved with our apparatus.

4 Random and Systematic Errors

Another point: The limitations of the technique of measurement may also give rise to the random errors. Imagine yourself measuring the period of a simple pendulum. You would like to start and stop your stopwatch when the bob is in an extreme position. However, due to the limitations of your reflexes, in a series of repetitions of the same measurement, you would start and stop your stopwatch sometimes earlier and sometimes later than the instant at which the bob reaches its extreme position. This will create a random error in your measurement.

Random errors are not the only source of errors in a meaurement. There are another kind of errors called the systematic errors. Technically, the systematic errors can be defined as those, which do not reveal themselves by merely repeating the measurement under identical conditions. The most fundamental error of this type is the 'least count' of your apparatus. Thus, if you measure a length using a ruler having a minimum marking of 1 mm your measurement will be uncertain by at least $\pm$ 0.5 mm and this uncertainty will persist in all measurements you make with the same ruler. A more serious source of systematic errors is the miscalibration of the instruments. For example, if your stop watch runs slow by one in thirty seconds, you will consistently under-estimate the period of a simple pendulum in every measurement with the same stop watch. Similarly a stretched (shrunk) ruler will always under-estimate (over estimate) any length being measured. Such errors can only be eliminated by careful calibration of all the measuring apparatus. In what follows we shall assume that all the systematic and human errors have been reduced so as to be negligible as compared to the required precision. Thus we are left with only the random errors.

In order to illustrate what has been said upto now, let us take an example. Suppose you want to measure the width of a slit of a mass spectrograph, approximately 1 mm wide. If you use a meter scale with least count 1 mm, the minimum error in your measurement will be $\pm$ 0.5 mm i.e. the error is of the order of the slit width. Note that this is the systematic error. You can improve upon this by using a scale with smaller graduation but you can't go beyond an error like 0.2 mm because of the limitations on the resolution of your eyes. Here you will get random errors due to the process of alignment of the scale graduations and the edges of the slit. You can do much better by mounting the slit on a travelling micoscope which can reduce the overall error to $\pm$ 0.01 mm. If an accuracy better than $10^{-2}mm$ is desired, you can use a laser interferometer (if you can afford it!) which can give you accuracies like $5 \times 10^{-8}mm$. When you reach this kind of an accuracy, (five parts in 10^8) the thermal noise of your apparatus (here the laser interferometer) becomes relevant. Due to the continuous thermal motion of the atoms emitting the laser radiation, the radiation is spread over a small wavelength interval $\Delta\lambda$ which [†] will limit

[†]Apart from the spread due to thermal motion, called Doppler width, there is an inherent width due

the accuracy with which the given length can be measured. The quantitative relationship between this accuracy and $\Delta\lambda$ depends on the details of the experiment.[‡]

5 Histograms (Frequency distributions)

Suppose we have a set of N values of a continuous variable x generated by N repetitions of the measurement of x (say $x_1, x_2, ... x_N$). We divide the range of values into intervals and count the number of values in each interval.

For example, a set of 10 values (all in same units)

$$26.4,\ 23.9,\ 25.1,\ 24.6,\ 22.7,\ 23.8,\ 25.1,\ 23.9,\ 25.3,\ 25.4$$

will give rise to the following table

Intervals(K)	$22-23$	$23-24$	$24-25$	$25-26$	$26-27$	$27-28$
Observations in an interval (n_k)	1	3	1	4	1	0

If a measurement falls exactly on the boundary between two intervals, we can assign half a measurement to each of the two.

The results such as those in the above table can be plotted in the form of a histogram as shown in fig. 1. The area of the rectangle drawn above each interval is the same as the fraction of measurements that fall in the interval. Thus if the width of the k^{th} interval is Δ_k then the height f_k of the rectangle drawn above the k^{th} interval is given by

$$\begin{aligned} f_k\Delta_k &= \text{fraction of measurements in the } k^{th} \text{ interval} \\ &= n_k/N \end{aligned}$$

Note that

$$\sum_k f_k\Delta_k = 1 \tag{3}$$

Check the above remarks using fig. 1.

Note that you have to choose Δ_k such that there are several readings in each of a large fraction of the number of intervals. If the intervals are made much too wide, then all (or almost all) the readings will fall in one interval. If the intervals are made too narrow then few of them will contain more than one reading and the histogram will be a large number

to quantum effects, called the natural line width. However, the Doppler width, whenever present, always masks the natural line width.

[‡]See for example, Optical Interferometery, by M.Francon, Academic Press, NY, Chapter 15.

of narrow rectangles almost all of the same height. When N is small this optimum size of Δ_k is relatively large but as N gets larger this optimum size of Δ_k gets smaller and smaller.

Out of the set of N values that we have we can construct two parameters which are of importance to us. First is the mean value of x :

$$\overline{x} \;=\; \sum_{i=1}^{N} x_i/N \;=\; \sum_{k} x_k f_k \Delta_k \tag{4}$$

In the latter sum k runs over intervals and x_k is the midpoint of the k^{th} interval.
The standard deviation, the measure of deviation from the mean is given by

$$\sigma_x \;=\; \sqrt{\frac{1}{N-1}\;\sum_{i=1}^{N}\;(x_i-\overline{x})^2}$$

If N is sufficiently large, the factor $1/(N-1)$ can be replaced by $1/N$. Note that ${\sigma_x}^2$ is the average value of the quantity $(x-\overline{x})^2$. The quantity ${\sigma_x}^2$ is called the variance and is denoted by $Var(x)$.

6 The Limiting Distribution

Almost all the measurements possess the following important property. As the number of measurements approaches infinity their histogram approaches some definite continuous curve described by a definite continuous function say $p(x)$. When this happens, the continuous function $p(x)$ (and the corresponding curve) is called the limiting distribution. Note that the limiting distribution cannot be exactly produced in practice, because this would require *infinitely* many repetitions of the measurement of the quantity x, under identical conditions. However, there is overwhelming evidence that for almost all measurements there exists a limiting distribution which our histogram approaches more and more closely as we make more and more repetitions of a measurement.

Since the limiting distribution is a limiting form of the histogram (in the limit $N \to \infty$) the following interpretations are easily understood. $p(x)dx =$ the fraction of measurements falling between x and $x+dx$; $\int_a^b p(x)dx =$ fraction of measurements falling between $x=a$ and $x=b$. If in addition, we impose the requirement (called the normalization condition) that

$$\int_{-\infty}^{\infty} p(x)dx \;=\; 1 \tag{5}$$

then the function $p(x)$ is called the probability density function with the interpretation

$p(x)dx$ = probability that any one measurement gives a value between x & $x+dx$

$\int_a^b p(x)dx$ = probability that any one measurement gives a value between $x=a$ & $x=b$

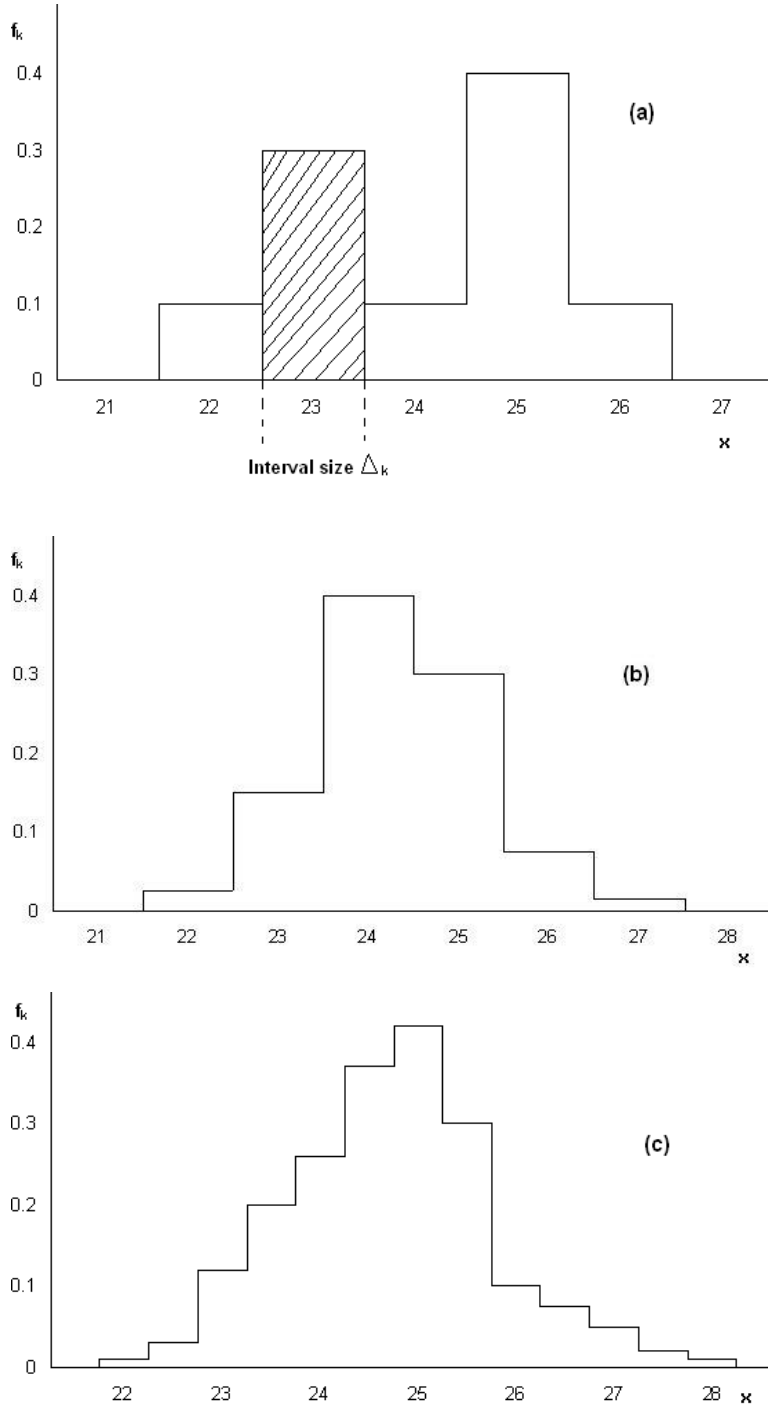

Figure 1: The histogram showing the fraction of measurements of x that fall in corresponding intervals for the number of values = (a) 10, (b) 100 and (c) 1000 respectively.

We have arrived at the following very important conclusion:

If we know (by some means) the normalised limiting distribution $p(x)$ for the measurement of a given quantity x with a given apparatus then we would know the probability of obtaining a value of x in any interval $a < x < b$.

If the measurement is very precise (with low overall uncertainty) then all the values obtained will be very close to the actual value of x (x_{true}). Furthermore, if all the errors are random, after many measurements, there will be as many observations above x_{true} as there are below it.

Under such conditions we expect our limiting distribution $p(x)$ to be

i centered on the actual value of x ($x = x_{true}$)

ii symmetric around $x = x_{true}$ i.e.

$$p(x_{true} + x) \;=\; p(x_{true} - x) \text{ for all } x$$

iii sharply peaked at $x = x_{true}$ i.e. $p(x)$ rapidly tends to zero as x goes away from $x = x_{true}$ on either side.

The last point does not stop us from integrating $p(x)$ from $-\infty$ to $+\infty$, because $p(x)$ is a well defined and continuous function of x over the whole range $-\infty$ to $+\infty$. A problem may arise if x cannot take negative values even in principle (e.g. kinetic energy, counts from a radioactive source etc.) However, in such cases x_{true} is generally found to be sufficiently away from $x = 0$ so that $p(x) \approx 0$ at $x = 0$. We can, then formally extrapolate $p(x)$ upto $-\infty$ because the fictitious contribution to $p(x)$ generated by such an extrapolation, namely $\int_{-\infty}^{0} p(x)dx$ is utterly negligible as compared to the total probability $\int_{-\infty}^{\infty} p(x)dx = 1$

7 The True Value Revisted

Let us elaborate on point(i) in the previous article. If we repeat the measurement of the quantity x under identical conditions, with negligible systematic errors, the resulting distribution will approximate the limiting distribution $p(x)$. This means that the centre of this distribution is very close to the centre of $p(x)$ i.e. x_{true}. But we can improve upon our distribution by performing more and more repetitions, each time getting better and better approximation to $p(x)$. Thus the centre of this distribution will tend to x_{true} as the number of repetitions tends to infinity. Thus, we can identify x_{true} as that value to which one approaches closer and closer as one makes more and more repetitions with negligible systematic errors. You may see that this identification of the true value was

really necessary. We had defined x_{true} to be the value of the physical quantity x, possessed by the system under observation at the time of measurement. However, due to the inevitable random errors, no measurement can exactly determine the true value of any continuous variable (a length, a time etc.) Thus the very existence of the true value can be questioned. For most of the physical quantities it is legitimate to assume that the true values exist and can be realised according to our new identification of the true value. There are situations, both pertaining to the physical quantities and to the systems under observation where there is no meaningful analogue of the true value. We shall discuss these later.

8 Mean and Standard deviation for $p(x)$

We can easily generalise the definitions of the mean and the variance of x for the limiting distribution $p(x)$.

We have

$$\overline{x} \;=\; \lim_{k\to\infty}\lim_{\Delta_k\to 0}\sum_k x_k f_k \Delta_k \qquad \text{with} \qquad \lim_{k\to\infty}\lim_{\Delta_k\to 0}\sum_k f_k\Delta_k \;=\; 1$$

giving

$$\overline{x} \;=\; \int_{-\infty}^{\infty} x\; p(x)dx \tag{6}$$

with

$$\int_{-\infty}^{\infty} p(x)dx \;=\; 1$$

Since ${\sigma_x}^2$ is the mean of the quantity $(x-\overline{x})^2$,

$${\sigma_x}^2 \;=\; \int_{-\infty}^{\infty} \left[(x-\overline{x})^2\; p(x)\right]\, dx$$

and the standard deviation

$$\sigma_x \;=\; \left[\int_{-\infty}^{\infty} \left[(x-\overline{x})^2\; p(x)\right]\, dx\right]^{\frac{1}{2}} \tag{7}$$

9 The Normal (Gaussian) Distribution

It is worthwhile to ask the following question: Is is possible to predict the form of the limiting distribution (i.e. of the function $p(x)$) for a given type of measurement? If this

can be accomplished we may be able to obtain the best estimate of the true value and also of the uncertainty contained in the value observed as a result of any single measurement, on the basis of the observed set of N values.

The answer to the above question is in the affirmative. This can be seen as follows:

Consider the measurement of a quantity x pertaining to a physical system whose true value exists in the sense defined in article 7 (we call such measurements to be deterministic or classical).

For such a measurement, the observed deviation from the true value x_{true} is a superposition of a very large (ideally infinite) number of fluctuations due to random vibrations of atoms in the apparatus, and many other sources of random errors due to the limitations of the measurement technique. We further assume that each of these fluctuations is very small (ideally infinitesimal) in magnitude and all these fluctuations are statistically independent i.e. nothing can be inferred about the magnitude of any one of them by using the knowledge of the magnitude of any other. Thus, the observed limiting distribution is the probability distribution for the linear superposition of a very large number of statistically independent fluctuations, the magnitude of every fluctuation being very small. Under these circumstances the observed limiting distribution is found to have the form which can be reliably approximated by the normalised Gaussian distribution,

$$P_{\mathsf{x},\sigma}(x) \;=\; (2\pi\sigma^2)^{-\frac{1}{2}}\; exp\;[-(x-\mathsf{x})^2/2\sigma^2] \tag{8}$$

You can check that this distribution is centered at $x = \mathsf{x}$ and is symmetric about it. Thus we identify x to be the true value of x.

The parameter σ is called the width parameter. Larger the σ the larger is the FWHM and vice-a-versa.

Let us find out the mean value of x viz. $\overline{x}$ for the Normal distribution.
We have

$$\begin{aligned}\overline{x} \;&=\; \int_{-\infty}^{\infty} x\; p(x)dx \\ &=\; (2\pi\sigma^2)^{-\frac{1}{2}}\; \int_{-\infty}^{\infty} x\; exp\;[-(x-\mathsf{x})^2/2\sigma^2]\; dx\end{aligned}$$

Put $y = x - \mathsf{x}$ and $dy = dx$. There is no change in the limits of integration.

$$\overline{x} \;=\; (2\pi\sigma^2)^{-\frac{1}{2}}\; \left(\int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy \;+\; \mathsf{x} \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} dy \right)$$

The first integrand is an odd function integrated over $-\infty$ to ∞ and hence the integral identically vanishes. The value of the second integral is $(2\pi\sigma^2)^{\frac{1}{2}}$ so that

$$\overline{x} \;=\; \mathsf{x} \tag{9}$$

Thus, if the limiting distribution is Gaussian then the mean value of x, after infinitely large number of trials, will be infinitely close to the true value x .

Since the variance ${\sigma_x}^2$ is the mean of the quantity $(x-\overline{x})^2$ we have

$${\sigma_x}^2 \;=\; \int_{-\infty}^{\infty} (x-\overline{x})^2 \; P_{\mathsf{x},\sigma}(x)\; dx$$

We leave it to the reader to evaluate this integral. The result is

$${\sigma_x}^2 \;=\; \sigma^2$$

Thus after infinitely large number of repetitions the observed standard deviation σ_x will be infinitely close to the width parameter of the Gaussian distribution. Henceforth, we shall call σ the standard deviation of the Gaussian distribution.

10 Interpretation of σ as the Measure of Uncertainty

It is quite legitimate to think of σ as the measure of uncertainty contained in the result of a single measurement of x. If you look at the curve of $P_{\mathsf{x},\sigma}(x)$ for a given σ you will notice that the probability of the values of x such that

$$|x-\mathsf{x}| > \sigma$$

falls off to zero quite rapidly i.e. the larger part (68%,see below) of the total area under the curve for $P_{\mathsf{x},\sigma}(x)$ lies over the interval $\mathsf{x}-\sigma$ to $\mathsf{x}+\sigma$. This area is nothing else but the probability that any one measurement will produce a value between $\mathsf{x}-\sigma$ and $\mathsf{x}+\sigma$:

$$P(\mathsf{x}-\sigma \le x \le \mathsf{x}+\sigma) \;=\; \int_{\mathsf{x}-\sigma}^{\mathsf{x}+\sigma} P_{\mathsf{x},\sigma}(x)\; dx$$

In fact if t is a real variable one can calculate

$$\begin{aligned} P(\mathsf{x}-t\sigma \le x \le \mathsf{x}+t\sigma) \;&=\; \int_{\mathsf{x}-t\sigma}^{\mathsf{x}+t\sigma} P_{\mathsf{x},\sigma}(x)\; dx \\ &=\; \int_{\mathsf{x}-t\sigma}^{\mathsf{x}+t\sigma} (2\pi\sigma^2)^{-\frac{1}{2}}\; exp\;[-(x-\mathsf{x})^2/2\sigma^2]\; dx \end{aligned}$$

Put $z = (x - \mathsf{x})/\sigma$ so that

$$P(x \text{ within } t\sigma) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-z^2/2}\, dz \tag{10}$$

This is the well known error function $\mathsf{erf}(t)$. In particular for $t = 1$, we get p (x within σ) = 68%

We can interpret our result in the following manner. If the limiting distribution is Gaussian with the standard deviation σ then we can say with 68% confidence that the result of a single measurement will lie within $\mathsf{x} \pm \sigma$ where x is the true value of x. This means that we can be 68% confident that the maximum uncertainty involved in a single measurement is $\pm\sigma$ i.e. if x_0 is the observed value,

$$x_0 - \sigma \leq x_{true} = \mathsf{x} \leq x_0 + \sigma \tag{11}$$

Another interpretation is that if the value of the standard deviation of the limiting gaussian distribution is known then out of any number of repetitions 68% will give a value lying between $\mathsf{x} \pm \sigma$.

11 Evaluation of the Best Estimate for x & σ

You may feel that our problem is now solved because if the limiting distribution is Gaussian then the value of the parameter x is infinitely close to the true value of x and the value of the parameter σ gives us the estimate (with corresponding confidence limits) of the uncertainty in a single measurement. How do we determine the values of x & σ? Note that these values will be determined by the given combination of the apparatus, the system under measurement and other external conditions. If this combination is changed the value of x & σ will, in general, change. Given a combination of the apparatus and the system under measurement, the only way to obtain the values of x & σ pertaining to the corresponding Gaussian distribution is to repeat the measurements, under identical conditions infinite number of times, generating infinite amount of the observed data. Thus, the problem now is to obtain the best estimates of the quantities x & σ using the finite amount of data we have.

Suppose we have the set of values $x_1, x_2, ... \; x_i ... \; x_N$ produced by N repetitions. We want to construct an expression for the probability of obtaining this set of values in N measurements, say $P(x_1, x_2, ... \; x_i ... \; x_N)$. Now all these N measurements are statistically independent because the probabilities of various values of x in one measurement are not affected by any other measurement. This statistical independence ensures that

$$P(x_1, x_2, ... \; x_i ..... \; x_N) \;=\; P(x_1)\; P(x_2) .... \; P(x_i) ..... \; P(x_N) \tag{12}$$

Since the probabilities of various values of x in any single measurement are normally distributed,

$$P(x_i) \;\propto\; \frac{1}{\sigma}\; exp\; [-(x_i - \mathsf{x})^2/2\sigma^2] \qquad i = 1, 2, ...N$$

$$P_{\mathsf{x},\sigma}(x_1, x_2, ... \; x_i ..... \; x_N) \;\propto\; \frac{1}{\sigma^N}\; exp\; \Big[- \sum (x_i - \mathsf{x})^2/2\sigma^2 \Big] \tag{13}$$

Note that here the set $\{x_i\}$ is *known* and the values of the parameters x & σ are *not known*. We want to obtain the best estimates of x & σ. We do this by using the principle of maximum likelihood, which can be stated as follows.

Given $x_1, x_2, ... \; x_N$, the best estimates for x & σ are those values, for which the probability $P_{\mathsf{x},\sigma}(x_1, x_2, ... \; x_N)$ is maximum. Obviously this means that the best estimate of x is that value for which

$$\sum_{i=1}^{N} (x_i - \mathsf{x})^2 \text{ is minimum}$$

Differentiate with reference to x and set the derivative equal to zero. This gives

$$\sum_{i=1}^{N} (x_i - \mathsf{x}) \;=\; 0$$

or

$$\mathsf{x} \;=\; \sum_{i=1}^{N} x_i/N \;=\; \overline{x} \tag{14}$$

Thus the mean of the N observations $(x_1, x_2, ... \; x_N)$ is the best estimate of x i.e. the true value of x that you can construct on the basis of these N observations. Of course, larger the number of observations (N) better is the 'best estimate' $\overline{x}$ of the true value x. The same is true for the best estimate of σ defined below.

To obtain the best estimate of the standard deviation of the limiting Gaussian distribution, on the basis of N observed values $(x_1, x_2, ... \; x_i ... \; x_N)$, we have to differentiate $P_{\mathsf{x},\sigma}(x_1, x_2, ... \; x_i ..... \; x_N)$ with reference to σ and equate the derivative to zero. This is slightly more complicated but straight forward and we leave the details to you. We find the following expression for the best estimate of σ

$$\sigma_x \;=\; \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mathsf{x})^2} \tag{15}$$

The true value x is unknown and therefore has to be replaced by its best estimate $\overline{x}$. Thus,

$$\sigma_x \;=\; \sqrt{\frac{1}{N}\;\sum_{i=1}^{N}\;(x_i-\overline{x})^2} \tag{16}$$

Now, the sum $(x_i-\mathsf{x})^2$ is a minimum when $\mathsf{x}=\overline{x}$ as we have already seen. Thus by replacing x by $\overline{x}$ as above, we have consistently underestimated the width σ. As a result it turns out that the best appoximation for σ is not eqn. 16 itself, but is obtained by multiplying it by the factor $\sqrt{\frac{N}{N-1}}$. Thus we finally obtain the best estimate for the width parameter to be

$$\sigma_x \;=\; \sqrt{\frac{1}{N-1}\;\sum_{i=1}^{N}\;(x_i-\overline{x})^2} \tag{17}$$

12 Two Important Questions

Let us pause and review a little. First, if the measurement of the quantity x is classical or deterministic *and* is subject only to the random errors (i.e. all systematic errors are negligible) then the corresponding limiting distribution is, to a very good approximation, the Normal distribution, centered at the true value of $x=\mathsf{x}$ and with width σ. The width is the 68% confidence limit, i.e. there is a 68% probability that any measurement will fall within a distance σ from the true value x. In practice neither x nor σ is known. What is known are our N measured values $x_1, x_2, ... \; x_N$ where N is as large as our time and patience allow us to make it. Based on these N values, our best estimate of the true value x is $\overline{x}=\sum x_i/N$ and our best estimate of the width σ is the standard deviation calculated from the set of values $\{x_i\}$ (eqn. 17)

Two important questions arise:

i What confidence do we have in our best estimate of the true value ($x=\mathsf{x}$) calculated from a single set of measured values $(x_1, x_2, ... \; x_N)$? That is, can we estimate the probability that the value of $\overline{x}$, calculated from a single set of values $\{x_i\}$ will fall within some distance, say $\sigma_{\overline{x}}$ from the true value x ? In particular we want to find out $\sigma_{\overline{x}}$ for which the above mentioned probability is 68% (68% confidence limit).

It is obvious that $\sigma_{\overline{x}}$ so defined, is the correct estimate of the uncertainty in our measurement, because it is the amount by which the best estimate of the true value is uncertain.

ii We have asserted that, under stipulated conditions, the limiting distribution is,

to a good approximation, Gaussian. Now the question is: How good is this approximation? Since all that we have is the observed set of N values, the correct question to be asked is: If we try to fit our observed data (i.e. the observed frequency distribution or the histogram) to the Normal distribution then how good would be this fit? We want to establish a quantitative measure for this 'goodness of fit'.

13 How to Report your Results

Before trying to answer any of the above questions let us state something which is really obvious, but has to be explicitly stated. On the basis of the N observed values $(x_1, x_2, ... x_N)$ how will you report your results in the form

$$(\text{value of } x) \;=\; x_{best} \pm \delta x \;?$$

We already know that for given $\{x_i\}$

$$x_{best} \;=\; \overline{x} \;=\; \sum x_i/N$$

δx must be the uncertainty in $\overline{x}$ i.e. $\sigma_{\overline{x}}$ as defined above.
Thus we have

$$(\text{value of } x) \;=\; \overline{x} \pm \sigma_{\overline{x}} \tag{18}$$

Now, suppose, you make N measurements of g, the acceleration due to gravity, on the basis of which you get these two values

$$g = 9.82 \; m/sec^2 \qquad \sigma_g = \pm 0.05 \; m/sec^2$$

Note that, it is meaningless to report the result of a single measurement as

$$g = 9.81268712 \; m/sec^2$$

This can be written

$$g = 9.81 + 0.00268712 \; m/sec^2$$

The second number on the RHS is less than σ_g and therefore could be replaced by any other number less than σ_g because your apparatus is not capable of differentiating between two values of g which differ by an amount less than σ_g. Thus, if you want to find out g from the measured values of the period and length of a simple pendulum, do not blindly copy down the 8 digit number displayed on your pocket calculator.

We make the above argument more rigorous in Appendix

14 The Propagation of Errors

The above example, of the determination of g using simple pendulum, opens up an interesting question. We have to measure the length and the period of the pendulum and then determine g through the relation

$$g \;=\; 2\pi\sqrt{\frac{l}{T^2}}$$

Thus g is not directly measured but is calculated from the values of two independently measured quantities, namely, the length and the period of the pendulum. These two measurements will have their own uncertainties, σ_l & σ_T. The question is, what is the consequent uncertainty in g?

Stated in more general terms, it means: Consider a set of physical quantities which can be directly measured. Let these quantities be denoted as $x, y, z, ...$ Let the limiting distribution fo each of these be normal, centered around the true value $X, Y, Z, ...$ and the width parameters be $\sigma_x, \sigma_y, \sigma_z, ...$ These σ s can be regarded as the measures of uncertainties for $x, y, z, ...$ Let $q(x, y, z, ...)$ be a continuous function of $x, y, z, ...$ The problem is to find out the probability distribution for the values of q and the uncertainty involved.

Assume that $\sigma_x, \sigma_y, \sigma_z, ...$ are fairly small so that the measured values of $x, y, z, ...$ are always close to their true values $X, Y, Z, ...$ In this case, the Taylor series expansion of $g(x, y, z, ...)$ around $X, Y, Z, ...$ can be expressed only in the first order terms in $x, y, z, ...$ Therefore

$$q(x,y,z,....) \;=\; q(X,Y,Z,....) \;+\; \Big(\frac{\partial q}{\partial x}\Big)_{X,Y,Z,..}(x-X) \;+\; \Big(\frac{\partial q}{\partial y}\Big)_{X,Y,Z,..}(y-Y) \;+$$
$$+\; \Big(\frac{\partial q}{\partial z}\Big)_{X,Y,Z,..}(z-Z) \;+\; \cdots$$

Note that $q(X, Y, Z, ...)$ and the partial derivatives evaluated at $X, Y, Z, ...$ are constants. One can verify in general that the probability distribution of the above linear combination of the variables $x, y, z, ...$ (which are individually distributed normally) is again a normal distribution with width parameter

$$\sigma_q \;=\; \sqrt{\Big(\frac{\partial q}{\partial x}\Big|_{X,Y,Z,..}\sigma_x\Big)^2 \;+\; \Big(\frac{\partial q}{\partial y}\Big|_{X,Y,Z,..}\sigma_y\Big)^2 \;+\; \Big(\frac{\partial q}{\partial z}\Big|_{X,Y,Z,..}\sigma_z\Big)^2 \;+\; \cdots} \qquad (19)$$

Obviously, σ_q is the required measure of the uncertainty involved in the value of q. The distribution is centered at $q(X, Y, Z, ...)$ Hence the probability that any value of q will lie within $q(X, Y, Z, ...) \pm \sigma_q$ is 68% (see article 10)

Eqn. 19 is called the formula for the propagation of errors.

Two simple illustrations:

1. $q(x,y) = x + y,\ q$ is centered around $X + Y$

$$\sigma_q \ = \ \sqrt{{\sigma_x}^2 + {\sigma_y}^2} \tag{20}$$

   This is generally referred to as addition in quadrature.

2. $q(x,y) = xy,\ q$ is centered around XY

$$\begin{aligned} \sigma_q \ &= \ \sqrt{Y^2\ {\sigma_x}^2 + X^2\ {\sigma_y}^2} \\ &= \ XY\ \sqrt{\left(\frac{\sigma_x}{X}\right)^2 + \left(\frac{\sigma_y}{Y}\right)^2} \end{aligned} \tag{21}$$

$$\text{(value of } q) \ = \ XY \pm XY\ \sqrt{\left(\frac{\sigma_x}{X}\right)^2 + \left(\frac{\sigma_y}{Y}\right)^2} \ = \ XY\ \left[1 \pm \sqrt{\left(\frac{\sigma_x}{X}\right)^2 + \left(\frac{\sigma_y}{Y}\right)^2}\right]$$

In practice X, Y, σ_x, σ_y are to be replaced by their best estimates.

We call $\sigma_x/\overline{x}$ & $\sigma_y/\overline{y}$ the fractional uncertainties in x & y.

15 Standard Deviation of the Mean (SDOM)

We now try to answer the first question raised in the article 12. It is necessary to know the reliability of the average of the N measurements. To do this, we must know the probability distribution of $\overline{x}$. We construct many sets of measurements under identical conditions each containing N values.
For each set we construct

$$\overline{x} \ = \ \frac{x_1 + x_2 + x_3 + ...x_N}{N}$$

Now we can regard x_1, x_2 etc. as variables spanning all the first, second etc. values in each set. Each one of these variables x_1, x_2 etc. is normally distributed. $\overline{x}$ can be looked upon as a function of these variables $x_1, x_2, ...$ etc. Hence we can apply the considerations of the article 14. Each of the variables is identically distributed with the same centre x and s.d. σ_x. Therefore the distribution of $\overline{x}$ would be centered at

$$\frac{\mathsf{x} + \mathsf{x} + \mathsf{x} + ...N \text{ times}}{N} \ = \ \mathsf{x}$$

The s.d of $\overline{x}$ will be according to the formula 19

$$\sigma_{\overline{x}} \;=\; \sqrt{\Big(\frac{\partial \overline{x}}{\partial x_1}\;\sigma_{x_1}\Big)^2 \;+\; \Big(\frac{\partial \overline{x}}{\partial x_2}\;\sigma_{x_2}\Big)^2 \;+\; \ldots\ldots \;+\; \Big(\frac{\partial \overline{x}}{\partial x_N}\;\sigma_{x_N}\Big)^2}$$

Since

$$\sigma_{x_1} = \sigma_{x_2} = \cdots \sigma_{x_N} = \sigma_x \qquad \& \qquad \frac{\partial \overline{x}}{\partial x_1} = \frac{\partial \overline{x}}{\partial x_2} = \cdots \frac{\partial \overline{x}}{\partial x_N} = \frac{1}{N}$$

We have

$$\sigma_{\overline{x}} \;=\; \frac{\sigma_x}{\sqrt{N}} \tag{22}$$

Thus $\overline{x}$ is found to be normally distributed, centered around the true value x and width $\sigma_{\overline{x}} = \sigma_x/\sqrt{N}$

Hence, by applying the same considerations as in article 10, we can say with 68% confidence that a value of $\overline{x}$ computed from a single set of N observations will lie between $\mathsf{x} \pm \sigma_{\overline{x}}$ i.e. $\mathsf{x} \pm \sigma_x/\sqrt{N}$. Note that $\sigma_{\overline{x}}$ as defined above is the width parameter of the probability distribution of $\overline{x}$. The best estimate of $\sigma_{\overline{x}}$ on the basis of a set of N repetitions is

$$\sigma_{\overline{x}} \;=\; \frac{1}{\sqrt{N}}\;\sqrt{\frac{1}{N-1}\;\sum_{i=1}^{N}\;(x_i - \overline{x})^2} \qquad\qquad \text{(best estimate)}$$

Henceforth, by $\sigma_{\overline{x}}$ we will understand the best estimate defined above. Thus, from a set of N repetitions for a physical quantity x we can report the value of x as

$$(\text{value of } x) \;=\; \overline{x} \;\pm\; \sigma_{\overline{x}}$$

which is the same as eqn. 18. We call $\sigma_{\overline{x}}$ to be the uncertainty in the mean, or the standard deviation of the mean (SDOM).

The factor $1/\sqrt{N}$ in $\sigma_{\overline{x}}$ gives an obvious way to improve the precision of our measurement. As we increase N the $\sigma_{\overline{x}}$ gets reduced by the factor $1/\sqrt{N}$. Thus the uncertainty in our best estimate, $\overline{x}$, gets reduced as we increase N. The probability distribution for $\overline{x}$ gets more and more peaked at the true value x.

Unfortunately, the factor $\sqrt{N}$ grows rather slowly as we increase N. Thus, if we wish to improve our precision by a factor of 10, just by increasing the number of repetitions N, we will have to increase N by a factor of 100 which is utterly impracticable. Furthermore, we are neglecting systematic errors which are not reduced by increasing the number of repetitions. Thus in practice, if you want to increase your precision appreciably, you will probably do better if you improve your technique (i.e. reduce the systematic

errors still further with the same method of measurement, or employ a totally different method of measurement which is more precise) than to rely merely on increased number of repetitions.

16 Confidence

In many situations we would like to compare the results of our measurement with some known, expected answer. For example, in an experiment to check conservation of momentum we may measure initial and final (total) momenta p & p' and would compare (within uncertainties) the value of $p - p'$ with the expected answer, zero. Any experiment where one measures a quantity for which an accurate accepted value is known (like g, the acceleration due to gravity) is also of this type.

We expect that if our measurement was precise enough, the observed value should lie close to the expected value. The question is: What is the maximum amount of discrepancy that can be tolerated between the observed and the expected value? Or, when can you say that the agreement between the observed and the expected value is satisfactory?

Let x be such a quantity and on the basis of N repetitions, the observed value of x is

$$(\text{value of } x) \;=\; \overline{x} \;\pm\; \sigma_{\overline{x}}$$

Let x_{exp} be the expected value of x. We define the discrepancy to be $\mid \overline{x} - x_{exp} \mid$. We assume that our measurement was governed by a normal distribution. We make two working hypotheses for this distribution.

1. The distribution is centered on x_{exp} i.e. x_{exp} is the true value of x, and in our measurement, all systematic errors were negligible.

2. The width parameter of the distribution *is equal to* $\sigma_{\overline{x}}$ (which is actually the best estimate of the width parameter). This approximation is reasonable if N is large.

We now want to decide whether our value $\overline{x}$ obtained is reasonable assuming our hypotheses to be correct (i.e. if the discrepancy $\mid \overline{x} - x_{exp} \mid$ is acceptable). If the answer is yes, we take our hypotheses to be correct; if the answer is no, then we doubt our hypotheses, and we must examine the possibilities of mistake in the measurement or calculations, undetected systematic errors, and the correctness of the expected answer, x_{exp}.

In the light of the discussion in the article 10 it is obvious that the discrepancy $\mid \overline{x} - x_{exp} \mid$ is acceptable if

$$\mid \overline{x} - x_{exp} \mid \;\leq\; \sigma_{\overline{x}} \tag{23}$$

Thus we draw a line of demarcation at $\sigma_{\overline{x}}$. Note that this line of demarcation is not a rigid one and, it depends on the particular situation. People relax it upto almost $2\sigma_{\overline{x}}$ i.e. in certain situations a discrepancy

$$\mid \overline{x} - x_{exp} \mid \ \leq \ 2\sigma_{\overline{x}}$$

is also acceptable.

It is convenient to express these inequalities in terms of probabilities. To do this, we define a parameter t through the equation

$$t\sigma_{\overline{x}} \ = \ \mid \overline{x} - x_{exp} \mid \tag{24}$$

We now ask for the probability of finding a value of $\overline{x}$ which differs from x_{exp} by $t\sigma_{\overline{x}}$ or more. This is

$$P(\text{outside } t\sigma_{\overline{x}}) \ = \ 1 - P(\text{within } t\sigma_{\overline{x}})$$

The probability on the RHS is just the normal error integral (see article 10) which is tabulated. Now, suppose $P(\text{outside } t\sigma_{\overline{x}}) \geq 32\%$, i.e. $P(\text{within } t\sigma_{\overline{x}}) \leq 68\%$. From article 10 this means that

$$t\sigma_{\overline{x}} \ = \ \mid \overline{x} - x_{exp} \mid \ \leq \ \sigma_{\overline{x}}$$

Thus the condition 23 is equivalent to

$$P(\text{outside } t\sigma_{\overline{x}}) \ \geq \ 32\%$$

Or in other words the discrepancy is acceptable if the above condition is satisfied.

In most of the above cases people agree to accept the discrepancy if

$$P(\text{outside } t\sigma_{\overline{x}}) \ > \ 5\%$$

or equivalently

$$\mid x - x_{exp} \mid \ < \ 1.96 \ \sigma_{\overline{x}}$$

We say that the discrepancy is acceptable within the 5% level of confidence.

17 The Weighted Average

Suppose that two laboratories A & B measure the same physical quantity x pertaining to similar physical systems (see example 4 in article 1). For the value of x they report

$$\begin{aligned} x_A \ &\pm \ \sigma_A \qquad &(\text{Lab } A) \\ x_B \ &\pm \ \sigma_B \qquad &(\text{Lab } B) \end{aligned}$$

Note that the two sets of repetitions, one at lab A and the other at lab B may not be under identical conditions. Infact the two labs might have employed entirely different methods of measurement. Thus, the sources of random errors at two places are not in general identical. Consequently, their precisions will differ and in general

$$\sigma_A \neq \sigma_B$$

The question is how to combine these two results to obtain a single best estimate of the true value of x. Obviously, we cannot just take the union of the two sets of repetitions, one at each lab and compute $\overline{x}$ & $\sigma_{\overline{x}}$

Assuming normal distribution for both the measurements we can write down expressions which are proportional to the probabilities of lab A obtaining the value x_A and lab B the value x_B.

$$P_x(x_A) \propto \frac{1}{\sigma_A} exp\ [-(x_A - \mathsf{x})^2/2{\sigma_A}^2]$$

$$P_x(x_B) \propto \frac{1}{\sigma_B} exp\ [-(x_B - \mathsf{x})^2/2{\sigma_B}^2]$$

where x is the unknown true value of x.
Since the two measurements are independent, the probability of lab A finding the value x_A *and* lab B the value x_B is

$$P_x(x_A, x_B) = P_x(x_A)\ P_x(x_B) \propto \frac{1}{(\sigma_A \sigma_B)}\ exp\ (-\chi^2/2)$$

where

$$\chi^2 = \left(\frac{x_A - \mathsf{x}}{\sigma_A}\right)^2 + \left(\frac{x_B - \mathsf{x}}{\sigma_B}\right)^2$$

Now, our principle of maximum likelihood asserts that the best estimate of the true value, x, is that value for which $P_x(x_A, x_B)$ is maximum, i.e. for which χ^2 is minimum.

Differentiating χ^2 w.r.t to x and equating to zero we have

$$\frac{x_A - \mathsf{x}}{{\sigma_A}^2} + \frac{x_B - \mathsf{x}}{{\sigma_B}^2} = 0$$

The solution of this eq$^{\text{n}}$ for x is the required best estimate

$$x_{best} = \frac{x_A}{{\sigma_A}^2} + \frac{x_B}{{\sigma_B}^2} \Big/ (1/{\sigma_A}^2 + 1/{\sigma_A}^2) \tag{25}$$

Putting $w_A = 1/{\sigma_A}^2$ & $w_B = 1/{\sigma_B}^2$

$$x_{best} = \frac{w_A x_A + w_B x_B}{w_A + w_B} \tag{26}$$

We call w_A & w_B the weights and x_{best} the weighted average of x_A & x_B.

Generalising, suppose we have n separate measurements of the same quantity x

$$x_1 \pm \sigma_1,\ x_2 \pm \sigma_2, ... \ x_n \pm \sigma_n$$

with their corresponding uncertainties $\sigma_1, \sigma_2, ... \ \sigma_n$, then the best estimate of the true value of x is

$$x_{best} \ = \ \sum_i^n w_i \ x_i \ \Big/ \ \sum_i^n w_i \tag{27}$$

where

$$w_i \ = \ 1/{\sigma_i}^2 \qquad i \ = \ 1, 2, ... \ n \tag{28}$$

are the weights.

Since x_{best} in eqn. 27 is a simple function of $x_1, x_2, ... \ x_n$, each of which is normally distributed, we can apply the considerations of article 14 to find out the uncertainty in x_{best} of eqn. 27

$$\sigma_{best} \ = \ \Big(\sum_i^n w_i \Big)^{-1/2} \tag{29}$$

18 Rejection of Data (Chauvenet's criterion)

Suppose we repeat the measurement of the quantity x, N times to get a set of values $x_1, x_2, ... \ x_N$. Let $\overline{x}$ & σ_x be the mean and standard deviation computed from this set of values. We assume that these values are distributed normally with the centre and the width given by $\overline{x}$ & σ_x respectively, given above. Suppose there is a particular value x_i such that $\mid x_i - \overline{x} \mid$ is very large so that x_i is highly improbable and we want to reject it. We say that a value so improbable must be the result of some mistake in that particular measurement and so it should be dropped. We now want to establish a quantitative criterion for rejecting such data values.

We first calculate

$$t_i \ = \ \frac{\mid x_i - \overline{x} \mid}{\sigma_x}$$

the number of standard deviations by which x_i differs from $\overline{x}$. Next we find the probability P (outside $t_i \sigma_x$) that a legitimate measurement will differ from $\overline{x}$ by t_i or more standard deviations.

$$P(\text{outside } t_i\sigma_x \ ) \ = \ 1 - P(\text{outside } t_i\sigma_x \ )$$

where RHS is tabulated.

Finally we multiply by N, the total number of repetitions, to arrive at

$$n \text{ (worse than } x_i \text{)} \;=\; N\; P(\text{outside } t_i\sigma_x)$$

This n is the number of measurements expected to be at least as bad as x_i.

Now the Chauvenet's criterion states that the concerned observation should be rejected if

$$n < 1/2$$

and be accepted otherwise.

After rejecting any measurements that fail Chauvenet's criterion, one recalculates $\overline{x}$ & σ_x from the remaining data. The resulting value of σ_x will be smaller than the original one and it may happen that, with the new σ_x, some more measurements will fail Chauvenet's criterion. However, it is generally agreed that Chauvenet's criterion should not be applied a second time using the recalculated values of $\overline{x}$ & σ_x.

19 The Poisson process

Upto now we have dealt with only the deterministic measurements. In this article we want to study a special case of measurements where the process under observation is itself random i.e. probabilistic. We have seen one example of such a process in article 1 namely the Brownian motion. Here we consider a particular experiment namely, counting the number of particles emitted by a radioactive source over a finite interval of time using a Geiger counter.

Consider a radioactive source containing a large number ($\sim 10^{20}$) of parent nuclei, which decay to their daughter nuclei by emitting an electron or a positron. For a given radioactive nucleus, the process of emission is random i.e. the time at which this nucleus will disintegrate cannot be predicted (we shall not go into the reasons for this). Thus, one has to ask for the probability that a given single nucleus will decay in unit time. Since all the parent nuclei in the source are identical, this probability is the same for every nucleus. If we now multiply p by the total number of parent nuclei in the source we get the expected average rate of emission from the source, which we denote by λ. We assume that the time for which we count the emitted particles is negligible as compared to the half life of the source. Therefore depletion in the number of parent nuclei over the time of measurement is negligible. Hence, λ is independent of time over the period of measurement.

Suppose we start our counter at $t = 0$ with all the registers equal to zero. Then after a time t the reading on the counter shows the number of particles counted in time t, say

$x(t)$. Since individual disintegrations are random in time, the values taken by $x(t)$ are also random. Noting that $x(t)$ is the quantity that is being measured we must ask for the probabilities for $x(t)$ taking various positive integer values (Note: $x(t)$ can take only positive integral values). Thus, we want to find the probabilities

$$P(x(t) = n) \qquad\qquad n = 0, 1, 2, ...$$

Note that these probabilities will be a function of time. We denote by $P_n(t)$ the probability that n particles are counted in time t (i.e.$P_n(t) = P(x(t) = n)$).

Also note that $x(t)$ is not a simple function of time. If we plug in a particular value of t in $x(t)$ we get a random variable and not a fixed number. Thus with $t = 10$ seconds, $x(10\text{seconds})$ is the number of counts recorded upto 10 seconds, which is a random variable and can take any positive integer value. Thus $x(t)$ represents a random (or stochastic) process which evolves in time.

We say that a single elementary event has occured at time t if a new particle emitted by the source is recorded by the counter i.e. the number of counts is increased by one at that time. Thus $P_n(t)$ is the probability of n events in time t. Henceforth, we shall talk in terms of these events. We also treat λ to be the expected average counting rate.

We now make the following assumptions regarding the process.

1. The expected average counting rate is independent of time. Thus the expected number of counts after any time t is given by λt.

2. No two or more events can occur at the same instant of time (i.e. simultaneously).

3. Let k_1 & k_2 denote the number of elementary events and t_a & t_b be any two *non-overlapping* time intervals then

$$P(k_1 \text{ in } t_a \ \&\ k_2 \text{ in } t_b) \ = \ P(k_1 \text{ in } t_a)\ P(k_2 \text{ in } t_b) \tag{30}$$

   i.e. the two events in the brackets on the RHS of the above equation are statistically independent.

Consider a time interval of length Δt immediately after a time t. We ask for the probability that at least one elementary event ocurs in this interval. By assumption (3), this probability is independent of t, or of the number of events (say n) which occured upto t. Thus the only quantity on which this probability can depend is the length of the interval Δt. It must also depend (parametrically) on a characteristic of your radioactive source, which is independent of time. Otherwise this probability will depend only on Δt irrespective of the source. The only characteristic of the source which is independent of

time (by assumption (1)) is λ. Therefore, this probability is a function of Δt & λ, say $w(\Delta t, \lambda)$. This can be written as

$$w(\Delta t, \lambda) \;=\; w_1 + w_2 + w_3 + ... \;=\; \sum_1^\infty w_n(\Delta t, \lambda) \tag{31}$$

where w_n is the probability that exactly n events will occur in Δt.

In terms of probabilities, our assumption (2) means that the probabilities w_n with $n > 1$ vanish much faster than Δt itself as $\Delta t \to 0$. Mathematically we say that

$$w_n \;=\; O(\Delta t) \qquad n > 1 \tag{32}$$

where $O(\Delta t)$ is a function such that

$$\lim_{\Delta t \to 0} \frac{O(\Delta t)}{\Delta t} \;=\; 0$$

Thus if we choose t to be sufficiently short we can neglect $w_2,\ w_3$..... occuring on the RHS of eqn. 31 as compared to w_1. Thus we have

$$w(\Delta t, \lambda) \;=\; w_1(\Delta t, \lambda) \tag{33}$$

where $w_1(\Delta t, \lambda)$ is a function of Δt & λ which goes to zero as $\Delta t \to 0$ not faster than Δt itself. The most general expression satisfying this condition is $\lambda \Delta t$ i.e. w_1 is proportional to Δt with proportionality constant λ. Thus we have, the probability that exactly one event will occur in Δt is equal to $\lambda \Delta t$, provided Δt is sufficiently short. As we shall see later Δt must be short enough to satisfy $\lambda \Delta t \ll 1$.

The probability that no event takes place in Δt is obviously given by $(1 - \lambda \Delta t)$ under the approximation 33. Denote this by w_0.

Now we ask for the probability that n events will occur in time $t + \Delta t$ i.e. $P_n(t + \Delta t$, with $n > 0$. If this is short enough, there are only two ways in which n events could be recorded upto $t + \Delta t$.

i $n - 1$ events were recorded upto time t and one event was recorded in the interval Δt after time t.

ii n events were recorded upto time t and no event was recorded in the interval Δt after time t.

Therefore we have

$$\begin{aligned} P_n(t + \Delta t) \;&=\; P_{n-1}(t).w_1 \;+\; P_n(t) w_0 \\ &=\; P_{n-1}(t)\ \lambda \Delta t \;+\; P_n(t)\ (1 - \lambda \Delta t) \end{aligned} \tag{34}$$

It follows immediately from this equation that

$$\frac{dP_n(t)}{dt} \;=\; \lim_{\Delta t\to 0}\frac{P_n(t+\Delta t)\;-\;P_n(t)}{\Delta t} \;=\; \lambda[P_{n-1}(t)\;-\;P_n(t)] \qquad n>0 \qquad (35)$$

For the case $n = 0$ the first alternative above is to be dropped. Hence for $P_0(t)$ we have

$$\frac{dP_0(t)}{dt} \;=\; -\,\lambda P_0(t) \qquad n=0 \qquad (36)$$

Since at $t = 0$ the Geiger counter read zero the initial condition is

$$P_0(0) \;=\; 1 \qquad (37)$$

The set of differential equations (35) and (36) together with the initial condition (37) are sufficient to determine the probability distribution $P_n(t)$. Thus we can solve the equation for $n = 0$ (eqn. 36) immediately to give, using eqn. 37

$$P_0(t) \;=\; e^{-\lambda t}$$

Substituting this in eqn. 35 with $n = 1$ and solving gives

$$P_1(t) \;=\; \lambda t\; e^{-\lambda t}$$

Substituting $P_1(t)$ into eqn. 35 with $n = 2$ we get

$$P_2(t) \;=\; \frac{(\lambda t)^2}{2!}\; e^{-\lambda t}$$

Repeating this procedure for successive $P_n(t)$ we get

$$P_n(t) \;=\; e^{-\lambda t}\;(\lambda t)^n/n! \qquad\qquad n=0,1,2.... \qquad (38)$$

This probability distribution, which depends on the parameter λt is called the Poisson distribution.

Now the only missing link is how to find λt from the number of recorded events in time t , which is the only information you have. Since λ is the average rate of counting, λt may be found by recording the counts over time t many number of times and finding the average number of counts. We shall justify this procedure.

Suppose we repeatedly record the counts over a fixed period of time, say t. Let the number of repetitions be N. Let n be the number of counts recorded. Let,

$$\begin{aligned}
N_0 &: \text{ No. of experiments with } && n \;=\; 0\\
N_1 &: && \text{"} \qquad\quad n \;=\; 1\\
N_2 &: && \text{"} \qquad\quad n \;=\; 2\\
&\vdots\\
N_k &: && \text{"} \qquad\quad n \;=\; k\\
&\vdots
\end{aligned}$$

We have $N = N_0 + N_1 + N_2 + ... + N_k + ...$
Let T be the total number of counts recorded in N experiments

$$T = 0.N_0 + 1.N_1 + 2.N_2 + ... + k.N_k + ... \tag{39}$$

Now if our N is sufficiently large, then the probability that k counts will be recorded in any given experiment is very close to N_k/N. Thus

$$\begin{aligned} P_k(t) &= N_k/N \\ N_k &= N\ P_k(t) \qquad\qquad k = 0, 1, 2..... \end{aligned}$$

substituting this in eqn. 39 we have

$$\begin{aligned} T &= N[P_1(t) + 2P_2(t) + 3P_3(t) + ... + kP_k(t) + ...] \\ &= N\Big[\lambda\ te^{-\lambda t} + 2\frac{(\lambda t)^2}{2!}\ e^{-\lambda t} + ... + k\frac{(\lambda t)^k}{k!}\ e^{-\lambda t} + ...\Big] \\ &= N\lambda\ te^{-\lambda t}\Big[1 + \frac{(\lambda t)}{1!} + ... + \frac{(\lambda t)^{k-1}}{(k-1)!} + ...\Big] \\ &= N\lambda\ te^{-\lambda t}.\ e^{\lambda t}\ =\ N\lambda\ t \end{aligned}$$

$$\therefore \qquad \lambda t\ =\ T/N \tag{40}$$

which was the result to be proved.

20 Some Properties of the Poisson Distribution

We formally write our Poisson distribution as

$$P(n, \mu)\ =\ \frac{\mu^n}{n!}\ e^{-\mu} \tag{41}$$

In this notation our $P_n(t)$ will read $P(n, \lambda t)$.

Let us calculate the average value of n say $\overline{n}$

$$\begin{aligned} \overline{n} &= \sum_{n=0}^{\infty} n\ P(n, \mu)\ =\ \sum_{n=0}^{\infty} n\ e^{-\mu}\ \frac{\mu^n}{n!} \\ &= \mu e^{-\mu}\ \sum_{n=1}^{\infty} \frac{\mu^{n-1}}{(n-1)!}\ =\ \mu e^{-\mu}.\ e^{\mu}\ =\ \mu \end{aligned} \tag{42}$$

$\overline{n}$ is nothing else but the average number of events over a large number of repetitions of the underlying experiment. This justifies eqn. 41. We can easily show that $P(n,\mu)$ is normalised i.e.

$$\sum_{n=0}^{\infty} n\ P(n,\mu)\ =\ 1 \tag{43}$$

Let us calculate

$$\sigma_n{}^2\ =\ \overline{(n-\overline{n})^2}\ =\ \overline{n^2} - (\overline{n})^2$$

We already know $(\overline{n})^2 = \mu^2$

Similar calculation gives

$$\begin{aligned} \overline{n^2} &= \mu^2 + \mu \\ \therefore \quad \sigma_n{}^2 &= \mu \end{aligned}$$

or

$$\sigma_n\ =\ \sqrt{\mu} \tag{44}$$

Thus, the Poisson distribution with mean count μ has a standard deviation $\sqrt{\mu}$.

If we carry out our counting procedure once, say for time t, and get a reading n, then using the principle of maximum likelihood it is easily seen that our best estimate for the expected mean count $\mu(=\lambda t)$ is $\mu_{best} = n$. From eqn. 44 it follows that our best estimate for the standard deviation is just $\sqrt{n}$. Thus if we make a single counting experiment for a time t and get the result n, then our answer for the expected average count for that time interval ($\mu = \lambda t$) is

$$n \pm \sqrt{n} \tag{45}$$

In other words $\mu(=\lambda t)$ is expected to lie in the interval $n \pm \sqrt{n}$ on the basis of a single measurement. In this sense $\sqrt{n}$ is the uncertainty in $\mu(=\lambda t)$. If we would count for a longer time, we would get a large value of n which would mean a longer uncertainty $\sqrt{n}$. However, the fractional uncertainty which is

$$\text{Fractional uncertainty}\ =\ \frac{\sqrt{n}}{n}\ =\ \frac{1}{\sqrt{n}} \tag{46}$$

would *decrease* if we counted for a longer time.

We can make $\mu(=\lambda t)$ larger and larger by counting for larger and larger time and/or by increasing the strength of the source. In this case $n \gg 1$, (minimum increment in n is unity), and can be treated as a continuous variable. It can further be proved that as $\mu(=\lambda t) \to \infty$ the Poisson distribution $P(n,\mu)$ assumes a form of a Gaussian or Normal distribution with the same mean and variance, where n is a continuous, rather than a discrete variable (see Appendix.). Symbolically,

$$P(n,\mu) \approx P_{\mathrm{x},\sigma}(n) \qquad (\mu \text{ large}) \tag{47}$$

where $$\mathsf{x} = \mu \quad \& \quad \sigma = \sqrt{\mu}$$

The approximation 47 is called the Gaussian approximation to the Poisson distribution. This is an excellent approximation provided μ is large (in practice $\mu > 20$).

You should note very clearly, the difference between the notions of the Poisson process and the Poisson distribution. The Poisson process is a sequence of events occuring at random in time, satisfying certain assumptions. This is described by a random function like $x(t)$. The Poisson distribution is given by eqn. 41 which gives the probabilities for various values of a single random variable. Thus, $P_n(t)$ is the Poisson distribution for the random variable which gets generated when we plug in a particular value of t in $x(t)$.

21 Examples

The two most common examples of the Poisson process are counting the disintegrations of radioactive nuclei and counting the arrival of cosmic ray particles.

A very important example of the Poisson distribution is an experiment to study an expected limiting distribution, like the Normal distribution. The Normal distribution $P_{\mathsf{x},\sigma}(x)$ tells us how many measurements of a quantity x are expected to fall in the interval from $x = a$ & $x = b$. In practice, the observed number is seldom exactly the expected number. Thus we can treat the observed number of measurements lying between $x = a$ & $x = b$ as a random variable and ask for its probability distribution.[§] This probability distribution is Poisson, provided the interval (a, b) is not too large. Thus if the expected number of measurements is n, then the observed number can be expected to differ from n by a number of order $\sqrt{n}$.

An enormous number of stochastic processes in real life can be approximated by the Poisson process. You can discuss this point with your instructor.

22 The χ^2 test for the Goodness of fit

We shall now try to answer the second question posed in article 12. Let us restate the problem as follows:

Suppose we make N identical and independent repetitions of a measurement of a physical quantity x, giving us a set of N values. Our conjecture is that these values are

[§]The corresponding experiment is as follows: In a set of N repetitions, we count the number of results falling in the interval $a \leq x \leq b$, say ν. If we repeat this procedure many times over, we shall get a set of values for ν. This experiment is analogous to counting the particles emitted by a radioactive source for a *fixed* interval of time.

distributed according to a certain probability distribution, say Gaussian (whatever is said below is true for any probability distribution provided it has a smooth enough p.d.f.). We are obviously assuming that x is a continuous variable.

We make the following hypothesis.

The observed values of x are distributed according to the Gaussian distribution $P_{\mathsf{x},\sigma}(x)$ where

$$\mathsf{x} \;=\; \overline{x} \;=\; \sum_{i=1}^{N} x_i/N \qquad \& \qquad \sigma \;=\; \sigma_{best} \;=\; \sqrt{\frac{1}{N-1}\;\sum_{i=1}^{N}\;(x_i-\overline{x})^2} \tag{48}$$

We want to make a quantitative criterion to test whether our hypothesis is true or false. To do this, we will have to compare the probability that a measurement will result in a value lying in an interval $a < x < b$ with the observed frequency of occurence of values between $a < x < b$. Alternatively, after dividing the whole range of values of x into n intervals of width say σ or its multiples, we can compare the expected number of occurences in the k^{th} interval $a_k < x < a_{k+1}$, say E_k,

$$\begin{aligned} E_k \;&=\; N\; P(x \text{ within } k^{\text{th}} \text{ interval}) \\ &=\; N \int_{a_k}^{a_{k+1}} P_{\mathsf{x},\sigma}(x)\; dx \qquad k = 1, 2, 3, ...n \end{aligned}$$

with the observed number of values lying in the k^{th} interval, say $O_k (k = 1, 2, ...n)$.

We can construct the number

$$O_k - E_k \qquad\qquad k = 1, 2, ...n$$

If our hypothesis is true, we expect this number to be small for all k. The question is: when can we say that this number is small enough for our hypothesis to be true? To answer this question we must decide how large we would expect $(O_k - E_k)$ to be if our hypothesis was true. From what has been said in the last section 21, it follows that for every k, O_k obeys the Poisson distribution with mean E_k and standard deviation $\sqrt{E_k}$ Thus the two numbers to be compared are the deviation $(O_k - E_k)$ and the expected size of its fluctuation $\sqrt{E_k}$.

These considerations lead us to consider the ratio

$$\frac{O_k - E_k}{\sqrt{E_k}} \qquad\qquad k = 1, 2, 3, ...n \tag{49}$$

Now we can say that if our hypothesis is true, this number must be of the order of unity for most of the n intervals. However, the above expression can have both positive and negative values for any given k so that it is difficult to construct a single criterion, involving all the n intervals, to test our hypothesis. Hence, it is natural to square the number in eqn. 49 for each k and then sum over all the intervals $k = 1, 2.....n$. This defines a number called Chi-squared,

$$\chi^2 \;=\; \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k} \tag{50}$$

In general, the individual terms in the sum (eqn. 50) are expected to be of the order of unity (if our hypothesis is true) and there are n terms in the sum. Therefore if

$$\chi^2 \leq n \tag{51}$$

(i.e. χ^2 is of the order of n or less) the observed and the expected distributions agree about as well as could be expected. Thus in this case our hypothesis will be true. Otherwise (i.e. when $\chi^2 \gg n$) the hypothesis will be false and cannot be accepted.

23 The Lorentzian distribution; an example

We have already mentioned that the definition of χ^2 and its utility to test whether the observed data follows a particular distribution, is independent of the form of this distribution. Thus the χ^2 test is applicable even when the expected distribution is Poisson or Binomial, for example. In many atomic and nuclear experiments the expected distribution of the measured variable x (actually an energy) is the Lorentzian distribution,

$$p(x) \;=\; \frac{1}{(x - \mathsf{x})^2 + \gamma^2} \tag{52}$$

where x & γ are constants. Since any probability distribution is normalised, we have

$$P(a < x < b) \;=\; \int_a^b p(x)\; dx$$

so that, for the k^{th} interval running from $x = a_k$ to $x = a_{k+1}$, the expected number of values in the k^{th} interval, after N repetitions, is

$$\begin{aligned} E_k \;&=\; N\; P(a_k < x < a_{k+1}) \\ &=\; N \int_{a_k}^{a_{k+1}} p(x)\; dx \end{aligned} \tag{53}$$

where $p(x)$ is given by eqn. 52. Once the E_k's hace been found, the χ^2 can be defined in exactly the same way and the criterion 51 also remains the same.

When the measured variable x is a discrete variable, having finite range, we need not divide the range of x into intervals to find the E_k's. We can find the expected number of occurences for every value of x separately. In practice, it is found more convenient to divide the range of x into intervals, even if x is discrete.

24 The General Definition of χ^2

The χ^2 test has much wider applications than the one being considered here. In all cases χ^2 is the sum of squares with the general form

$$\chi^2 \;=\; \sum_1^n \left(\frac{\text{Observed value} - \text{Expected value}}{\text{Standard deviation}} \right)^2 \tag{54}$$

Also, χ^2 is an indicator of the agreement between the observed and the expected values of some variable. If the agreement is good $\chi^2 \leq n$ and if it is bad $\chi^2 \gg n$.

We can only use χ^2 to test this agreement if we know the expected values and the standard deviation and can therefore calculate (54). In our case, namely, a test of a distribution, these are known accurately enough, with E_k given by the distribution and the standard deviation given by $\sqrt{E_k}$. Another situation where χ^2 test can be applied is the measurement of two variables x & y where y is expected to be some definite function of x.

$$y \;=\; f(x) \tag{55}$$

(like $y = A + Bx$). Suppose, we have N measured pairs (x_i, y_i) where x_i have negligible uncertainty and y_i have known uncertainties σ_i. Here the expected values of y_i are $f(x_i)$ and we can test how well y fits the function $f(x)$ by calculating

$$\chi^2 \;=\; \sum_1^N \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 \tag{56}$$

Again the agreement is satisfactory if $\chi^2 \leq N$ and unsatisfactory if $\chi^2 \gg N$. Thus if $\chi^2 \leq N$, the hypothesis $y = f(x)$ is accepted, not otherwise.

25 Degrees of Freedom and Reduced χ^2

Through the hypothesis made in article 22 we have imposed two external conditions on the limiting distribution obeyed by the observed data namely $\mathsf{x} = \overline{x}$ & $\sigma = \sigma_{best}$. It is

easy to see that if these external conditions are to be true, all of the N values in the set cannot be independent of each other. Once any $N-2$ values are realised the remaining two values are fixed in terms of these $N-2$ values (through eqn. 48) by the requirement that the two external conditions (or constraints) be satisfied. We say that the number of degrees of freedom for the problem is $N-2$.

We shall adopt the following working definition for the number of degrees of freedom: The number of degrees of freedom d in a statistical calculation is defined as the number of observed data minus the number of parameters computed from the data and used in the calculation (which is the same as the number of constraints).

For the problem of the χ^2 test, the number of observed data are the number of observations O_k in n intervals, $k = 1, 2, ...n$. Thus the number of observed data is just n, the number of intervals. Therefore for the problem at hand,

$$d \;=\; n - c \tag{57}$$

where n is the number of intervals and c is the number of parameters that had to be calculated from the data in order to compute the expected numbers E_k (i.e. c is the number of constraints).

In the χ^2 test considered here the number of constraints is three. They are given by the equations

$$\begin{aligned}
\mathsf{x} \;&=\; \overline{x} \;=\; \sum_{i=1}^{N} x_i/N \\
\sigma \;&=\; \sigma_{best} \;=\; \sqrt{\frac{1}{N-1}\;\sum_{i=1}^{N}\;(x_i-\overline{x})^2} \\
N \;&=\; \sum_{k=1}^{n} O_k
\end{aligned}$$

Thus the number of degrees of freedom

$$d \;=\; n - 3 \tag{58}$$

where n is the number of intervals.

Now, it can be shown (though we shall not do so here) that the expected value of χ^2 is precisely d, the number of degrees of freedom. This means that if we could repeat the whole set of N measurements infinitely many times and compute χ^2 each time, then the average of these values of χ^2 is d. Thus, even for a single measurement, it is more

appropriate to compare the value of χ^2 with d, rather than with n, so as to accept or reject our hypothesis.

A slightly more convenient procedure is to introduce a reduced χ squared, denoted by $\widetilde{\chi}^{\,2}$ and defined as

$$\widetilde{\chi}^{\,2} \;=\; \chi^2/d \tag{59}$$

Since the expected value of χ^2 is d,

$$(\text{expected value of } \widetilde{\chi}^{\,2}) \;=\; 1 \tag{60}$$

Thus our test can be stated as follows:

$$\text{If } \widetilde{\chi}^{\,2} \leq 1 \tag{61}$$

we accept the hypothesis that our data was governed by the expected distribution, and if

$$\widetilde{\chi}^{\,2} \gg 1 \tag{62}$$

the hypothesis is rejected.

26 The probabilities for χ^2

Now we want to make our χ^2 test more quantitative. Suppose on the basis of N measurements we compute reduced chi squared, say $\widetilde{\chi}^{\,2} = 1.80$. Now the question is: Is a value $\widetilde{\chi}^{\,2} = 1.80$ sufficiently larger than one to rule out the expected Gaussian distribution or not?

To answer this question, we first suppose that our hypothesis is true i.e. our measurements were governed by the expected distribution (a Gaussian in this example). With this assumption, one can calculate the probability of obtaining a value of $\widetilde{\chi}^{\,2} \geq 1.80$. This probability turns out to be

$$P(\widetilde{\chi}^{\,2} \geq 1.80) \approx 18\% \tag{63}$$

Now we have to decide whether this probability is unreasonably small (so that if our hypothesis is true, it is extrememly unlikely to obtain a value of $\widetilde{\chi}^{\,2} \geq 1.80$) and therefore our hypothesis has to be rejected. The question of drawing this line of demarcation at a value, say a, such that $P(\widetilde{\chi}^{\,2} \geq 1.80) \;\geq\; a$ implies the acceptance and $P(\widetilde{\chi}^{\,2} \geq 1.80) \;<\; a$ implies the rejection of our hypothesis, does not have a unique answer. Rather, it is dictated by the nature of the problem at hand. However the value $a = 5\%$ is generally accepted.

Summarising, we first compute the value $\widetilde{\chi}^{\,2} = \widetilde{\chi}_0^{\,2}$ on the basis of the set of N observed values. Then assuming our hypothesis that the observed data are governed by the expected probability distribution to be true, we find out $P(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2})$.

If
$$P(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2}) \;\geq\; 5\% \tag{64}$$

we accept the hypothesis, otherwise reject it. Some people choose to relax this condition to

$$P(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2}) \;\geq\; 1\% \tag{65}$$

In the first case we say that the hypothesis is rejected at 5% significance level and in the second case at 1% significance level.

The calculation of the probabilities $P(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2})$ is beyond the scope of this note. However, these probabilities are tabulated. It turns out that the probability of getting any particular value of $\widetilde{\chi}^{\,2}$ depends on the number of degrees of freedom. Thus we could write the probability of interest as $P_d(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2})$ to emphasise its dependence on d.

The ususal calculation of $P_d(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2})$ treats the observed numbers O_k as continuous variables, distributed according to the Gaussian distribution around mean E_k and with standard deviation $\sqrt{E_k}$. This is obviously the Gaussian approximation to the Poisson distribution which is justified if all the numbers (O_k & E_k) are reasonably large ($\gg 1$). Under these conditions the tabulated probabilities $P_d(\widetilde{\chi}^{\,2} \geq \widetilde{\chi}_0^{\,2})$ can safely be used. This means that the intervals must be so chosen that the expected number of counts in each interval is reasonably large (at least 5 or so). For the same reason, the number of intervals should not be too small to contradict the assumption that O_k is continuous.

27 Limitations of the classical approach

We want to build an argument to show that the classical approach or the deterministic type of measurements, with which we have been basically concerned in this note, are a kind of approximation. In th classical situation, every system is in a definite state of motion and the corresponding values of the physical quanitites (the true values) can be approached continually as we repeat their measurements more and more number of times, with negligible systematic errors.

Whenever a measurement of any variable is performed, the system concerned is allowed to interact with the apparatus, and the resulting change in the state of the apparatus (e.g. change in the pointer of an ammeter) is related to the value of that variable at that time. As a result of this interaction, the state of the system under observation must also change. Therefore the result of our measurement is uncertain by the amount by which the value of the variable being measured is changed, due to the process of measurement

itself. It is easily seen that it is impossible to quantify this uncertainty. The only way to do so is to try to measure the change due to the interaction of the system and the apparatus. This you can do only by making the first apparatus and the system interact with some other apparatus which will introduce its own uncertainty.

The only way out is to ask whether you can neglect this uncertainty altogether i.e. whether this uncertainty is utterly negligible compared to all the other random and systematic errors. For those situations in which this is true, the classical approach is valid.

It is found that, as the rest mass of the system under observation gets smaller and smaller (i.e. comparable to that of an atom, electron etc.) the uncertainty due to the process of measurement itself cannot be neglected. In such situations the classical approach fails completely and a qualitatively new theory is required to deal with these measurements and with the states of motion of such systems. This theory is quantum mechanics and the corresponding systems are called the quantum systems. The subject of quantum theory of measurement is very deep and totally out of the scope of this note.

Appendix I

In order to clarify the concepts regarding the inherent random errors due to the thermal noise of the apparatus, we consider the following problem. ¶

Suppose you want to measure a mechanical motion that you know is oscillating at a single frequency. The question is: What is the smallest vibration amplitude that can, in principle, be measured?

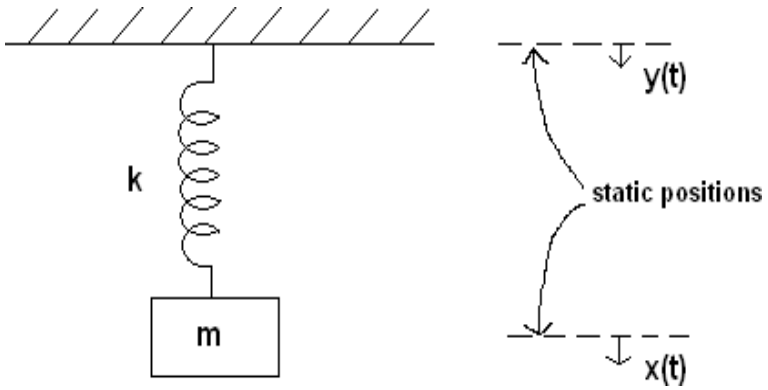

Figure 2:

As a model for a detector we note that if you hang a weight on a spring and then move the upper end of the spring up and down, the amplitude of the weight will be much larger than the driving amplitude if you are at the resonant frequency of the mass and spring assembly. It is essentially a mechanical amplifier and serves as a good candidate for a sensitive detector (fig. 2)

The equation of motion is

$$m\ddot{x} \;=\; -K(x-y) - D\dot{x} \tag{I.1}$$

where the overdot denotes time differentiation. We have assumed that the damping term, $D\dot{x}$, depends only on the motion of the mass (such as the air resistance). Another term, $D\dot{y}$, could be added and would take into account the losses in the spring but it would clutter up the following analysis without adding anything essential to it.
Suppose $y(t) = y_0\, e^{i\omega t}$. Let $x(t) = x_0\, e^{i\omega t}$, then

$$(-\omega^2 m + i\omega d + K)x_0 \;=\; K\; y_0$$

¶In this Appendix, we have reproduced, (almost verbatim) a part of the article 'The Theoretical Limit for the Detectability of a signal' by D.M. Zipoy in the 'Methods of Experimental physics', Volume 8.

or
$$x_0 = \frac{K\ y_0}{K - m\omega^2 + i\omega D} \tag{I.2}$$

This is a maximum when $\omega^2 = K/m \equiv \omega_0^2$ (for D small)

$$x_{0_{max}} \approx \frac{m\omega_0^2}{i\omega_0 D}\ y_0 \equiv -iQy_0 \quad \text{where } Q \equiv \frac{m\omega_0}{D} \tag{I.3}$$

In order to make x_0 large we want to make Q (the 'Quality' factor) large, that is, we want a large mass and low damping. Even at this point there are limits to how big Q can be. The mass has to be small enough so that it does not cause the support to come crashing down. It is conceivable that the damping D could be made exceedingly small by putting everything in a vacuum and perhaps cooling it to a very low temperature at which point the internal losses in the spring tend to become small. From a practical point of view there is a limit to how large you want Q to be. It is determined by your lifetime as an upper limit (we presume). The reason for this limit is the following. At some time or other the detector has to be constructed, hooked onto the support, and released. it takes a while before the oscillation of the mass builds up to its steady state amplitude, x_0; how long will this take? You can easily verify that the characteristic expoential buildup time is $2Q/\omega_0$. Now if the period of oscillation is $1sec$ and $Q = 10^7$, then the buildup time is about a month and so you would have to wait a few months after you released the mass before you could make a measurement.

Let us now go on to the main point of this problem. What is the thermal noise associated with this device?

The source of noise arises from the equipartition theorem of statistical mechanics, that innocuous theorem you heard in your elementary physics course and promptly forgot. It is used as follows. You know each atom in the mass, spring and support has $1/2kT$ of energy in each degree of freedom. That is, they are twitching around such that they have $1/2kT$ of energy in the "x component" of kinetic energy, etc. This concept can be generalised to say that each normal mode of a system constitutes a degree of freedom (actually two degrees of freedom, one for the kinetic and one for the potential energy of the mode of vibration). Remember that there are two ways (at least) of describing the motion of a blob of material; either in terms of the position of each atom as a function of time or by giving the amplitude and phases of all the possible normal modes of vibration. This latter description is most useful for our purposes because it enables us to immediately write down the mean square fluctuation in position of the mass m. (The lowest mode of this system is just what is usually thought of as the reasonant mode; there are other modes which involve transverse and longitudinal vibrations of the spring, for instance). In all there are $6N$ modes of the system (where N is the number of atoms of the mass, spring, and support). We will only be concerned with the ususal mode since the others

are ususally at a different enough frequency that they can be discriminated against. We have

$$\frac{1}{2}K < x^2 >_N \;=\; \frac{1}{2}kT \quad \text{or} \quad < x^2 >_N \;=\; \frac{kT}{m\omega_0^2} \tag{I.4}$$

The signal to noise ratio is:

$$\left\langle \frac{S}{N} \right\rangle \;=\; \frac{1/2 \mid x_0 \mid^2}{< x^2 >_N} \;=\; \frac{m\omega_0^2\, Q^2}{2kT}\, y_0^2 \tag{I.5}$$

For a numerical example, suppose we are interested in a frequency of $1Hz$. Let $m = 10\,g,\ \omega_0 = 2\pi,\ Q = 100$ (a reasonable value), $T = 300\,^\circ\mathrm{K}$, then $S/N \approx 10^{10} Y_0$. So with the above detector we could measure vibration amplitudes of about $10^{-10} cm$. This could be improved by increasing m and Q and decreasing the temperature.

Appendix II

We try to answer the following question: Having measured a quantity x with estimated error Δx (as in eqn. 18), upto how many decimals should the value of x be reported?

While writing the numeric value of the quantity x, we may do the following spurious variations.

1. We may, if possible, add zeroes (as many as we please) just after the decimal point and multiply the number by a suitable power of 10, thus spuriously changing the number of decimals.

2. We also might add as many trailing zeroes as we please (after the last non-zero decimal)

It is important to note that in both the above cases it is possible to get rid of these zeroes once and for all by multiplying by a suitable power of ten, if necessary. These zeroes do not convey any new information.

There is a third case also. Suppose the value of x is a five digit number (the integral part has 5 digits) but the last digit is unknown or ambiguous. In order to ensure that the number has five digits we put a zero in the fifth place. This zero cannot be dropped but it need not be counted as significant. In order to make the above points precise, we define what are called the signifcant figures.

Definition: A significant figure is any one of the digits $1, 2 - - - 9$; and 0 is a significant figure except when it is used to fix the decimal point or to fill the places of unknown or discarded digits.
Examples:

1. 0.00263 : significant figures are $2, 6, 3$ and hence the number of significant figures is 3.

2. 3809 : significant figures are $3, 8, 0, 9$ and hence the number of significant figures is 4.

Count the number of significant figures in the following numbers.

$$4.63 \times 10^4, \quad 4.630 \times 10^4, \quad 4.6300 \times 10^4$$

The question stated above can also be read as: Upto how many significant figures should the value of x be reported?

The Absolute and Relative errors

Definition: The *absolute error* of a number, a measurement or a computation is the absolute difference between the true value of the quantity and its approximate value as given, or obtained by measurement or computation i.e.

$$\text{Absolute error } E_A \;=\; \mid x_{true} - x_{observed} \mid$$

The *relative error* is the absolute error divided by the true value of the quantity.

$$E_R \;=\; E_A / x_{true}$$

Remarks:

1. Note that the error Δx in x, expressed as in eqn. 18 is the absolute error in x.

2. Relative errors are independent of the unit of measurement whereas the absolute errors are expressed in terms of the unit used.

3. If the value of a quantity is correct upto n significant figures it is obvious that its absolute error cannot be greater than half a unit in the n^{th} place. For example, if the value of the quantity is 4.629 and is correct to four significant figures, its absolute error is not greater than $0.001 \times \frac{1}{2} = 0.0005$.

4. The true index of the accuracy of a measurement or of a computation is the relative error. For example, if the width of a $1\,mm$ slit is measured to the nearest thousandth of an mm, the result is less accurate than the measurement of the earth-moon distance to the nearest kilometer. For although the absolute errors in two measurements are $5 \times 10^{-4}\,mm$ and $5 \times 10^{5}\,mm$ respectively, the relative errors are $(5 \times 10^{-4}\,mm / 1\,mm) = 5 \times 10^{-4}$ and $(1\,km / 384400\,km) \approx 2.6 \times 10^{-6}$. Hence in the measurement of the slit we make an error of one part in 2000 while in the case of the earth-moon distance measurement we make an error of one part in 384400. The latter measurement is clearly more accurate even if its *absolute error* is 10^9 times that of the former!

As a result of our measurements we get the estimates for the true value of the quantity being measured and the absolute error (random + systematic) involved. From these we can calculate the relative error. While reporting the value of the quantity, we want it to be consistent with the relative error, i.e. the value should contain the correct number of significant figures so that the value does not look more accurate than that is permitted by the relative error. If the value reported contains n significant figures and is consistent with the corresponding relative error then we say that the value is correct upto n significant figures. Thus in order to answer the question we have asked, we must establish a quantitative relationship between the relative error and the number of significant figures

allowed by this error. In what follows, we state two theorems (without proof) which do exactly this job.

Theorem I

If the first significant figure of the number (standing for the value to be reported) is k and if the number is correct to n significant figures, then the relative error is less than $1/(k \times 10^{n-1})$

Examples:

1. Let the number 864.32 be correct upto five significant figures. Then $k = 8$, $n = 5$ and $E_A \leq 0.01 \times \frac{1}{2} = 0.005$. One can easily verify that the relative error
$$E_R < \frac{1}{8 \times 10^4}$$
2. let the number 369230 be correct upto five significant figures. Then $k = 3$, $n = 5$ and $E_A \leq 10 \times \frac{1}{2} = 5$. Again one can see that
$$E_R < \frac{1}{3 \times 10^4}$$
3. Let the number 0.0800 be correct upto three significant figures. Show that the above theorem applies here.

Actually our problem is converse to the theorem stated. We want to find the number of correct figures corresponding to a given relative error. We cannot use the converse of theorem I because that is not true. In order to solve our problem we state the next theorem.

Theorem II

If the relative error in an approximate value is less than $1/((k+1) \times 10^{n-1})$, the value is correct to n significant figures or atleast is in error by less than a unit in the n^{th} significant figure.

Corollary 1: If $E_R < 1/(2(k+1) \times 10^{n-1})$ then the given value is correct to n significant figures.

Corollary 2: Since k may have any value from 1 to 9, it is evident that $k + 1$ may have any value from 2 to 10. Hence the upper and lower limits of the fraction $1/(2(k+1) \times 10^{n-1})$ are $1/(4 \times 10^{n-1})$ and $1/(2 \times 10^n)$ respectively. We can therefore assert that:

If the relative error of any value is not greater than $1/(2 \times 10^n)$, the value is certainly correct upto n significant figures.

Theorem II (with corollaries 1 & 2) clinches the issue.

Appendix III

Continuous Approximation to the Poisson Process

The Poisson distribution (eqn. 41) gives us probabilities for the various values taken by a discrete varaible (say number of counts in time t). As the mean of the distribution gets larger and larger the minimum increment in the value of the discrete variable can be taken to be infinitesimal and the discrete variable can be approximated by a continuous variable. In our problem, this means that, instead of asking for the probabilities $P_n(t)$, $n = 0, 1, 2....$, we now ask for the probabilities $P(x,t)\, 0 \leq x \leq \infty$, where x is a continuous real variable.

At the end of the article 20, we have said that, in the aove approximation, $P(x,t)$ is the p.d.f of the normal distribution. Let us now try to see how this happens.

We go back to the counting process under study. The rate equation for the probability $P_n(t)$, eqn. 35, now becomes

$$\frac{\partial P(x,t)}{\partial t} \;=\; \lambda\; P(x-1,t) \;-\; \lambda P(x,t) \tag{III.1}$$

If we now expand the first term in the R.H.S. of eqn. III.1 in a Taylor series about $p(x,t)$ we get

$$\frac{\partial P(x,t)}{\partial t} \;=\; \lambda\left[P(x,t) - \frac{\partial P(x,t)}{\partial x} + \frac{1}{2}\,\frac{\partial^2 P(x,t)}{\partial x^2}\right] \;-\; \lambda P(x,t)$$

Retaining only the terms of second order yields the equation

$$\frac{\partial P}{\partial t} \;=\; -\,\lambda\,\frac{\partial P}{\partial x} \;+\; \frac{1}{2}\,\lambda\,\frac{\partial^2 P}{\partial x^2} \tag{III.2}$$

Eqn. III.2 is called the diffusion type equation. we shall not attempt to solve it here. The solution is

$$P(x,t) \;=\; (2\pi\;\lambda t)^{-\frac{1}{2}}\; exp\left(-\frac{(x-\lambda t)^2}{2\lambda t}\right) \tag{III.3}$$

You should check whether eqn. III.3 is really a solution of (III.2). We can easily recognise (III.3) to be the normal p.d.f with mean and variance $\mu = \lambda t$.

Note that the p.d.f. given by eqn. III.3 will assign non-zero probabilities to the negative values of x which is unacceptable (why?). If the mean is large and positive, however, the probability that x takes a negative value is negligible. (Sketch the p.d.f. to understand this).

References:

There are numerous books dealing with the analysis of errors in physical measurements. Here are a few of them:

1. Statistical Methods in Measurements
   O.L. Lacy, Mac Millan, 1953.

2. Experimentation: An Introduction to Measurement Theory and Experiment Design
   D.C. Baird, Prentice Hall, 1962.

3. Experimental Measurements: Precision, Error and Truth
   N.C. Bardford, Addison-Wesley, 1967.

4. Statistical Treatment of Experimental Data
   H.D. Young, McGraw-Hill, 1962.

   Following books are more advanced in their mathematical level of treatment, and more complete in their coverage.

5. Data Reduction and Error Analysis for Physical Sciences
   P.R. Bevington, McGraw-Hill, 1969.

6. Statistical Treatment of Experimental Data
   S.L. Mever, John Wiley, 1975.

7. The Analysis of Physical Measurements
   E.M. Pugh and G.H. Winslow, Addison-Wesley, 1966.

8. Statistical Methods in Experimental Physics
   W.T. Eadie, D. Drijand, F.E. James, M. Roos & B. Sadoulet, North-Holland, 1971.

   A rather old but still very valuable reference is :

9. Calculus of Observations
   E. Whittaker and G. Robinson, Blackie and Son Ltd., 1952 (4th edition) esp. articles 83 to 87.

   The following two references are strongly recommended:

10. The Theoretical Limit for the Detectability of a Signal – in the Methods of Experimental Physics, Vol. 8 pp 1
    D.M.Zipoy, L. Marton & S.F. Honyark Ed.

11. Fluctuation Theory in Physical Measurements – in Reports on progress in Physics, Vol. XVI, pp. 266
C.W. Mccombie, 1953.

There are innumerable books on probability and its applications. However, a very standard reference on classical probability theory is,

12. An Introduction to the Probability Theory & its Applications
W. Feller, Wiley Easten Ltd., 1972.

There are, again, a large number of books dealing with stochastic processes in general, and the Poisson process in particular. Most of the commonly cited books are a difficult reading. A fairly complete and readable account of the Poisson process is given in

13. The Elements of Stochastic Processes
N.T.J. Bailey, John Wiley & Sons, 1964.

14. Probability, Random Variables, & Stochastic processes
A. Papoulis, McGraw-Hill, 1984.

The problem discussed in Appendix Appendix II comes under the general circle of numerical analysis. Again the number of books are large. Following are two of the standard references.

15. Numerical Mathematical Analysis
J.B. Scarborough, Oxford & IBM Publishing Co. (VII edition), 1971.

16. Introduction to Numerical Analysis (2^{nd} edition)
F.B. Hildebrand, McGraw-Hill, 1974.