Error Analysis and the Gaussian Distribution

In experimental science theory lives or dies based on the results of experimental evidence and thus the analysis of this evidence is a critical part of the scientific method. Imagine an experiment designed to measure the acceleration of gravity. To make the situation concrete suppose a pendulum is used to measure g, and suppose further that the period of the pendulum can be measured accurately to plus or minus 1 m sec and that the length of the pendulum is known to the nearest .1 mm. Using these uncertainties with what degree of precision can you determine g and what if any other conditions must be specified?

The question above can be answered provided the nature of the errors involved are well defined. Generally speaking errors can be divided into two major categories; random errors and systematic errors. Systematic errors are errors that cause a result to move away from the actual result in a single direction[1]. For example if the timer used to measure the period of the pendulum in the example cited above runs slow the period will be consistently underestimated and the resulting value of g will be higher than the true value. Systematic errors are often very difficult to identify, diabolical in their damage to experimental results, pathological in nature, and generally just plain ugly. Thus systematic errors are given lip service and only infrequently dealt with in introductory laboratory exercises. Systematic errors are often the result of improperly calibrated instruments and/or the consistent use of questionable or improper procedures in experimental technique.

Systematic errors, due to their stealth nature, frequently go unnoticed. An experimental result that is wrong for no apparent reason may be due to a systematic error. Finding a systematic error is aided by checking the calibration of all instruments and a careful examination of the experimental technique employed. Our primary focus will be on random errors and how to interpret their influence on the experimental results obtained. Systematic errors will be secondary but you should consider them in your error analysis.

Random errors are small errors that move in both directions and have a tendency to cancel one another, which makes them tractable using statistical analysis. Experience and intuition suggests that a better value of an unknown quantity can be obtained by measuring it several times and taking the average value to obtain an estimate of the actual or so called true value. Systematic errors may not be analyzed statistically.

Systematic and random errors have been defined in an intuitive way and both must be considered in a well-designed experiment. To further complicate the situation systematic and random errors occur together and the job of the experimenter is to recognize them and reduce the systematic errors to the point that they are consistent with the degree of precision the experimenter is attempting to obtain. For example in the problem specified above the method used to determine the time to within 1 m sec must be checked, otherwise the values for the period may produce systematic errors that push the values of g high or low without the experimenters knowledge.

For the moment it will be assumed that systematic errors have been reduced to a level that is small compared to the random errors. This means the instruments are calibrated and the experimental procedures employed are consistent with small systematic errors or small as compared to the random errors. In the example above suppose the experimenter finds the following values for the length of the pendulum:

99.37, 99.38, 99.38, 99.39, 99.39, 99.39, 99.39, 99.40, 99.41, 99.42 cm

[1] In some cases systematic errors may move in both directions; for example temperature variations over the course of an experiment may cause error to move in both directions. In many cases, however, the systematic error causes variations in only a single direction.

The data has been ordered so it is easy to see the frequency with each measurement occurred in the data set. From this data a histogram or frequency distribution could be constructed and the average value could be determined in the ordinary way or could alternately obtained using the frequency distribution, as we shall illustrate.

Let F be the frequency of a given measurement, for example for the data shown 99.38 occurs twice or with a frequency 2, and 99.39 occurs with a frequency 4. The average is just the sum divided by the total number or it can be written as:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{\sum_{j}^{M} F_j x_j}{N} = \frac{99.37 + 2\times 99.38 + 4\times 99.39....}{10}$$
$$\mu = \frac{\sum_{j}^{M} F_j x_j}{N} = \sum_{j}^{M} \frac{F_j}{N} x_j = \sum_{j}^{M} P_j x_j \tag{1.1}$$

In (1.1) the quantity P_j may be thought of as the probability of finding the corresponding x_j. This simple illustration suggests that measuring variables whose value is affected by random errors is a statistical process where P_j is the relative probability of obtaining the result x_j. Imagine a quantity you are trying to measure has a true value x_{true} and surrounding this true value are many imposters that are true value wannabes. Imagine all these unknowns being placed in a hat and the act of measuring is replaced by picking a value out of the hat. The distribution of wannabes depends upon the size of the random errors; for example if you are measuring a length and using a crude instrument the errors will be large but will be reduced in magnitude as your refine your measuring process by the use of a more accurate scale. The distribution of the samples will be characterized by the spread in values of the wannabes around the mean value, μ. The probability distribution for a random error that is as likely to move the value in either direction is called a Gaussian distribution. Such a distribution is characterized by two parameters, μ the mean or average value, and σ the standard deviation. In data analysis the Gaussian distribution is called the limiting distribution for the type of measurement we have described, however it is not the only limiting distribution that occurs in error analysis[2].

The probability density for the Gaussian distribution is

$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \tag{1.2}$$

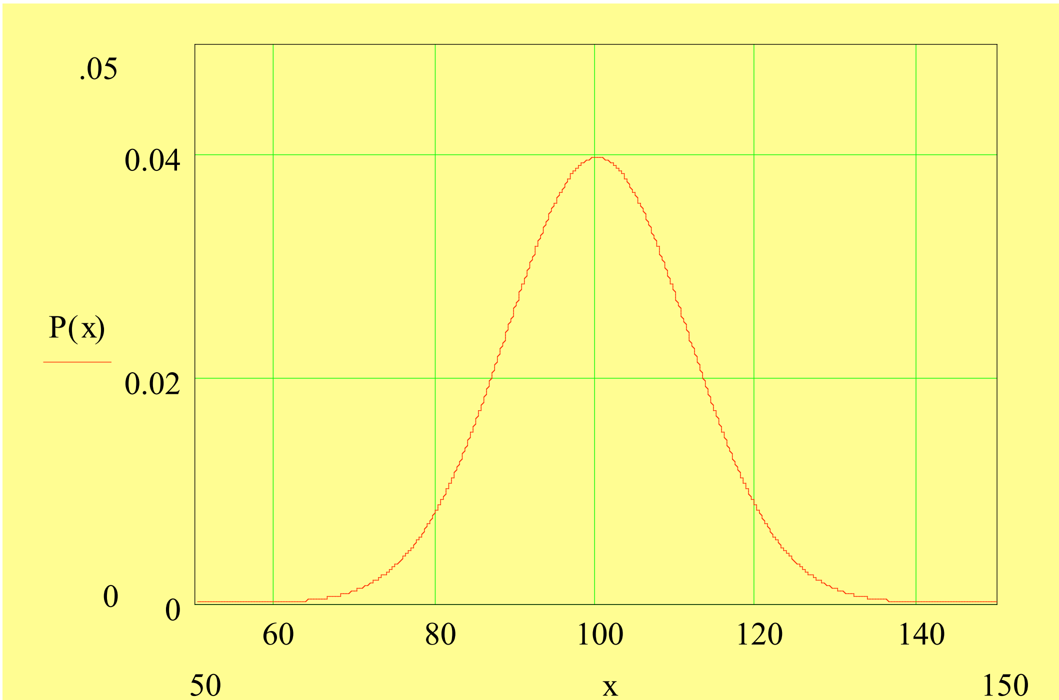
Equation (1.2) represents the probability that a given measurement will have a value x. The probability that the a measurement falls between x and x + dx is given as P(x)dx. In view of the fact that a measurement will have some value on the real number line it follows that

$$1 = \int_{-\infty}^{\infty} P(x)dx \tag{1.3}$$

There are a number of observations that can be made regarding (1.2); the function is symmetric with respect to μ = x and it decreases more rapidly with smaller σ. The standard deviation is a measure of

[2] Another example is the Poisson distribution, of which we shall say more later.

the precision of a measurement and is also proportional to the confidence of a given measurement. A perfect example to illustrate the Gaussian distribution is the so-called IQ test, which has been designed to have a mean value of 100 and a standard deviation of 16. (Note – there are many variations, however the one most often quoted is designed to have a mean of 100, which for such a test can be assigned.) The Gaussian distribution appears as:



Normal Distribution of IQ

The standard deviation is a measure of the sharpness of the curve; as σ increases the curve becomes flatter and as σ decreases the curve is sharper. Approximately 68 percent of all values lie between $\pm \sigma$, 95.4 % between $\pm 2\sigma$, and 99.7 between $\pm 3\sigma$.

It can be proven that if a measurement is subject to many small sources of random error and negligible systematic error, the values will be distributed around the mean, which is the best estimate of the true value. Additionally as the number of trials increases the range of uncertainty decreases as the square root of the number of datum. This latter quantity, namely σ/sqrt(N) is called the standard deviation of the mean (SDOM) or the standard error associated with N measurements as stipulated above.

Thus far we have discussed the normal or Gaussian distribution in general terms and have made an effort to convey the meaning of the probability distribution function as it applies to experimental measurement. At this point it is appropriate to put these concepts into more exact terms.

Gaussian Distribution - Quantitative

We begin by defining the quantities used to describe a measurement and show how they are related to one another and to the Gaussian distribution. In the course of what follows it will be assumed that systematic errors have been reduced to the point that they are smaller than the random errors and thus may be ignored.

Assume a set of data is obtained consisting of multiple measurements of some quantity, for example the length of the pendulum as stated above. As you know based on experience and intuition taking the average of this data will give you a better result than a single measurement – or at least that is probably what you assume. The good news is this is a correct assumption and is a property of random errors.

Random errors are errors that can push a result away from its true value in either the positive or negative direction and as such the random errors tend to cancel one another out. As a consequence of this canceling tendency as the number of measurement is increased the average more closely approaches the "true" value. The Greek letter mu, μ is used to define the mean or average value and is formally defined as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1.4}$$

The spread of values around the mean is as important as the mean value itself since the spread may be interpreted as a measure of the precision of a given measurement. As the precision is improved the spread of individual values about the mean becomes narrower. A quantity called the deviation is defined as:

$$\begin{aligned} &\delta_i = \mu - x_i \\ &but \\ &\sum_{i=1}^{N}\delta_i = \sum_{i=1}^{N}\mu - x_i = N\mu - \sum_{i=1}^{N} x_i = N\mu - N\mu \end{aligned} \tag{1.5}$$

Equation (1.5) shows that the sum of the deviations is zero. Make certain you understand the meaning of (1.5). Therefore using the sum of the deviations directly does not give us a useful measure of the deviation of the data about the mean. The variance, which is defined as the sum of the square deviations divided by the number of data is defined as:

$$\begin{aligned} &\text{Variance} = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(\mu - x_1)^2 \\ &\text{Standard Deviation} = \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mu - x_1)^2} \end{aligned} \tag{1.6}$$

There is a minor problem with the definition of the standard deviation as displayed in (1.6). Suppose you make one measurement, then μ = x_i and the standard deviation is zero, whereas if there are two different values then σ is suddenly not zero. It does not seem reasonable, and it can be shown that (later) a better definition of the standard deviation is given as:

$$\sigma = \sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}(\mu - x_1)^2} \tag{1.7}$$

It is customary to treat sigma as the uncertainty for a particular measurement. Indeed recall that approximately 68% of all measurements will lie within one standard deviation of the mean. This fact leads us to define the uncertainty of a measurement to be

$$\delta x = \sigma \tag{1.8}$$

There are instances when you might wish to modify (1.8) so that the uncertainty might be 1.2σ or maybe even 2σ if, for example, you wanted to increase your confidence level. In other words the definition in (1.8) is adopted as it is widely used.

As the number of observations increases the central or true value is given with greater accuracy – this means that as N increases we would expect sigma to decrease. In other words if we take many sets of measurements then we could calculate the standard deviation of the mean, or the mean of the means. Accordingly the error should be reduced and indeed this is exactly what does occur. The standard deviation of the mean is defined as

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} \tag{1.9}$$

The standard deviation decreases as one over the square root of the number of observations. This means we get better accuracy as the number of observations is increased, however it should be noted that to get a factor of ten increase in accuracy the number of observations has to increase 100 fold.

Justification

The Gaussian and Poisson[3] distributions can both be understood by looking at the binomial distribution as the former are both special cases of the latter. In order to introduce the topic the one dimensional "random walk" problem will be analyzed. To make this concrete consider a small object, say a feral or cat that can make steps of unity along the x-axis in either the positive or negative direction with equal probability. The position of said cat will then be x = ml, where m is the number of steps and l is the length of each step, and where m lies between –N and +N.

$$\begin{aligned} x &= ml \\ -N &\le m \le N \end{aligned} \tag{1.10}$$

l is redundant as we stated steps of unity, however we shall leave it in place should we wish to alter the length of the steps.

Let n_1 equal the number of steps to the right and n_2 the number of steps left so $n_1 + n_2 = N$. The net displacement measured to the right will then be $m = n_1 - n_2$. So

$$\begin{aligned} m &= n_1 - n_2 = n_1 - (N - n_1) \\ m &= 2n_1 - N \end{aligned} \tag{1.11}$$

Equation (1.11) shows that if N is odd then so is m and if N is even so is m. Now we assume the probability for a step right is equal to that of a step left. If p is the probability right and q is the probability left and if these are independent events the probability for a given situation is given as:

[3] The Poisson distribution is included and is well suited for other types of situations as will be discussed later.

$$P(n_1, n_2) = p^{n_1} q^{n_2} \tag{1.12}$$

However there are many ways to take n_1 steps right and n_2 left and the number is given as the number of ways n_1 can be taken without regard to order and likewise for n_2, which is given as:

$$\frac{N}{n_1! n_2!} \tag{1.13}$$

Exercise: Consider the case for N = 3. Show that there is only one way to end up at +3 or -3 and there are three ways to get to +1 or -1. Next show that these results are consistent with (1.13) and that they correspond to the coefficients a $(a+b)^3$. Try it for N = 4. Get the idea?

Now the probability of taking a total of N steps, n_1 right and n_2 left, is obtained by multiplying (1.12) times (1.13)

$$W_N(n_1) = \frac{N!}{n_1! n_2!} p^{n_1} q^{n_2} \tag{1.14}$$

This probability distribution is called binomial distribution. Recall the binomial expansion may be written as:

$$(p+q)^N = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n q^{N-n} \tag{1.15}$$

Note – The binomial expansion is very important and something you should know! The expression (1.15) is good for rational values of N as well as integer values.

Exercise: For p = 1 and q = x expand (1.15) for N = 3, N = -1, N = -1/2.

Thus the probability $P_N(m) = W_N(n_1)$. Now n_1 and n_2 can be expressed in terms of N and m as

$$\begin{aligned}
&n_1 = \frac{1}{2}(N+m) \ \ and \ \ n_2 = \frac{1}{2}(N-m) \\
&W_N(n_1) = \frac{N!}{n_1! n_2!} p^{n_1} q^{n_2} = \frac{N!}{[(N+m)/2]![(N-m)/2]!} p^{(N+m)/2} q^{(N-m)/2} \\
&P_N(m) = \frac{N!}{[(N+m)/2]![(N-m)/2]!} p^{(N+m)/2} q^{(N-m)/2} \qquad (1.16) \\
&for \ \ p = q = 1/2 \\
&P_N(m) = \frac{N!}{[(N+m)/2]![(N-m)/2]!} \left(\frac{1}{2}\right)^N
\end{aligned}$$

Exercise: Using Excel find the values of $P_N(m)$ for all possible values of m where $N = 20$. Plot the result as a bar graph. As you will see the shape of the curve is very similar to that of a Gaussian distribution. The binomial distribution is a discrete distribution and in the limiting case as N tends to infinity the binomial distribution becomes the Gaussian distribution. For additional detail the following link is of value: http://mathworld.wolfram.com/NormalDistribution.html ; as well as: http://www43.wolframalpha.com. Wolframalpha is a very useful tool for all kinds of stuff – Try it I think you will like it.