PHY224H1F/324H1S
Notes on Error Analysis

References:
J.R. Taylor: An Introduction to Error Analysis: The Study of Uncertainties in Physics Measurements, 2nd ed., University Science Books, 1997
P.R. Bevington, D.H. Robinson: Data Reduction and Error Analysis for the Physical Sciences, 3rd ed., McGraw Hill 2003

Introduction
Experimental errors are inevitable. In absolutely every scientific measurement there is a degree of uncertainty we usually cannot eliminate. Understanding errors and their implications is the only key to correctly estimate and minimize them.

In your first year of university physics you must have read a document on Error Analysis in Experimental Physical Sciences and eventually done all the related exercises and answered all the questions in the document. The document is located at:
http://www.upscale.utoronto.ca/PVB/Harrison/ErrorAnalysis/index.html
We strongly advised you to review it, including all the questions and exercises. We do this again!

1. Terms and definitions
The experimental error can be defined as: "difference between the observed value and the true value" (Merriam-Webster Dictionary). Uncertainties (errors) in experimental science can be separated into two categories: *random* and *systematic*.
Random errors fluctuate from one measurement to another. They may be due to: poor instrument sensitivity, random noise, random external disturbances, and statistical fluctuations (due to data sampling or counting).
Systematic errors usually shift measurements in a systematic way. They can be built into instruments. Systematic errors can be at least minimized by instrument calibration and appropriate use of equipment. Extraneous effects can also alter experimental results.
The terms *accuracy* and *precision* are often misused. *Experimental precision* means the degree of exactness of the experiment or how well the result has been obtained. Precision does not make reference to the true value; it is just a quality attribute. *Accuracy* refers to correctness and means how close the result is to the true value.
Accuracy depends on how well the systematic errors are compensated.
Precision depends on how well random errors are reduced.
Accuracy and precision must be taken into account simultaneously. All measurements are subject to both uncertainties.

2. One variable: the simple average and the standard deviation
Assume we want to measure a quantity x. We identified and reduced all systematic errors; we are left with only random uncertainties. We take N measurements of x. We know (from First Year Physics) that the best estimate for our measurements would be *the average (mean) value:*

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

The N values have been measured in the same conditions, by using the same equipment. However, they differ from each other and from the mean value because of the random uncertainties. In order to quantify these uncertainties, we define *the deviation* or *residual* of measurement *i,* from the mean value:

$$d_i = x_i - \bar{x} \qquad (2)$$

Very small residuals mean precise measurements. This suggests that residuals can be used to assess the reliability of measurements. One way of doing this assessment would be taking the average of all deviations. However, some deviations are positive and some are negative, so that the average would be zero.

Another procedure we may try is to take the average of the squares of all deviations, then to take the square root of the result. This 'root-mean-square' or 'RMS' approach has the advantage of yielding a final result with the same units as the measured values. The final number resulting from the RMS of deviations is called *the standard deviation* of measurements $x_1, x_2, ...x_N$:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad (3)$$

A better definition of the standard deviation would be:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad (3')$$

The difference between (3') and (3) is very small at large N values. At low N, expression from (3') calculates a slightly larger standard deviation than (3), but (3') is the only way of dealing with a very low number of measurements.

Standard deviation is the uncertainty to be used with any value from the measurement set $x_1, x_2, ...x_N$. To express the uncertainty in the mean value $\bar{x}$, we define *the standard deviation of the mean* or *the standard error:*

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} \qquad (4)$$

3. Distributions of values

Repeated measurements show a clear *distribution* of values around the mean. In order to manipulate and display the values, we have to know the properties of mathematical functions called *probability distributions.*

Three such distribution functions are important in data analysis:

- Gauss (normal) distribution
- Binomial distribution
- Poisson distribution

We shall cover in detail the Gaussian or normal error distribution since it is commonly used in analyzing experimental data. We will also discuss the Poisson distribution later.

In experiments characterized by N measurements of the same quantity, we can display data in the form of a histogram which has on the vertical the fraction F_i of the N measurements that gave the result x_i (where i = 1, 2, 3, ...N) and on the horizontal the measured values x_1, x_2, ...x_N.
As the number of measurements increases, the histogram changes into a quasi-continuous curve, close to a bell shape. The continuous curve is a graph of *the limiting distribution* and is described by a mathematical function called the normal distribution or Gauss function:

$$e^{-x^2/2\sigma^2} \qquad (5)$$

σ is a fixed parameter called *width* (we defined it before as standard deviation) .
Function from (5) is symmetrical about x = 0, is 1 at x = 0 and decreases to zero as $x \to \infty$.
A Gauss function centered on a point $X \neq 0$ would be expressed as:

$$e^{-(x-X)^2/2\sigma^2} \qquad (5')$$

The probability density P(x) is a very important quantity which defines the Gauss function. P(x)dx means the fraction of measurements that fall between x and x+dx or *the probability* that a measurement will fall between x and x+dx. $\int_{-\infty}^{+\infty} P(x)dx = 1$ is the normalization condition (total probability must be 1).
The Gauss function can be written and interpreted as a probability density if we arrange it to satisfy the normalization condition:

$$P(x) = N\, e^{-(x-X)^2/2\sigma^2} \qquad (6)$$

and $\int_{-\infty}^{\infty} P(x)dx = 1 = N\sigma\sqrt{2\pi}$ (6')

Formula (6') makes use of the fact that: $\int_{-\infty}^{+\infty} e^{-z^2} dz = \sqrt{\pi}$, which is proved in calculus courses. The normalization factor N is calculated and the final form of the Gauss or normal distribution is expressed as:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2} = P(x) \qquad (7)$$

X is the center of the distribution; σ is the width.
Calculation of the mean value follows as:

$$\bar{x} = \int_{-\infty}^{\infty} xG(x)dx = X \qquad (8)$$

Q1. Verify (8).

An interesting application would be to calculate σ and confirm its meaning as standard deviation.
Standard deviation defined in (3) and (4), is the average of the squared deviations $(x-\bar{x})^2$:

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 P(x)dx \qquad (9)$$

In evaluating the integral, substitute $\bar{x}$ by X and integrate by parts, obtaining:

$$\sigma_x^2 = \sigma^2 \qquad (10)$$

(10) proves that "the width" parameter of Gauss function (σ) and the standard deviation σ_x are the same.

4. One variable: the problem of the weighted average

Sometimes, we take several independent measurements of the same physical quantity. In order to express the result of our experimental work, we need to combine them into a best estimate of that quantity.
Suppose we have measured quantity x in two separate runs:

$$\begin{aligned} &x = x_A + \sigma_A \\ &and: \\ &x = x_B + \sigma_B \end{aligned} \qquad (11)$$

where: $x_{A,B}$ are the mean values from measurements A or B, $\sigma_{A,B}$ are the corresponding standard errors.
Assuming that the two measurements are *consistent*, defined by the following statement: $|x_A - x_B| \cong \sigma_A, \sigma_B$. We need to calculate the best estimate (true value) of variable x. We shall name the unknown true value of x by X.
Assuming that both measurements are governed by the Gauss distribution, the probability of obtaining the value x_A is approximated by:

$$P_X(x_A) \approx \frac{1}{\sigma_A} e^{-(x_A - X)^2 / 2\sigma_A^2} \qquad (12a)$$

Correspondingly, the probability of obtaining the value x_B is:

$$P_X(x_B) \approx \frac{1}{\sigma_B} e^{-(x_B - X)^2 / 2\sigma_B^2} \qquad (12b)$$

The probability of finding value A *and* value B is the product of probabilities (12a) and (12b):

$$P_X(x_A, x_B) = P_X(x_A)\, P_X(x_B) \approx \frac{1}{\sigma_A \sigma_B} e^{-\frac{\chi^2}{2}} \qquad (13)$$

The exponent, called **Chi-squared (χ^2)** is expressed as:

$$\chi^2 = \left(\frac{x_A - X}{\sigma_A}\right)^2 + \left(\frac{x_B - X}{\sigma_B}\right)^2 \qquad (14)$$

The principle of likelihood states that our best estimate for the unknown *true value X* is that value for which the actual data x_A and x_B are most likely to occur.
According to this, when reaching the best estimated value, the overall probability (13) has to be a maximum, or the value of chi-squared has to be a minimum. This method of finding the best estimate is often called **the method of least squares.**
To find X, we look for the minimum of χ^2 and we obtain:

$$X = \left(\frac{x_A}{\sigma_A^2} + \frac{x_B}{\sigma_B^2}\right) \Bigg/ \left(\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2}\right) \qquad (15)$$ This is *the best estimate for X*

We define the *weights* of x_A and x_B to be: $w_A = \frac{1}{\sigma_A^2}$ and $w_B = \frac{1}{\sigma_B^2}$

We can now introduce *the weighted averages* of x_A and x_B as:

$$x_{wav} = \frac{w_A x_A + w_B x_B}{w_A + w_B} \qquad (16)$$ This is also *the best estimate for X*

If the uncertainties of x_A and x_B are identical ($\sigma_A = \sigma_B$), (16) reduces to the simple average of x_A and x_B.

The best estimate (16) would be closer to x_A if $\sigma_A < \sigma_B$, which means measurement A would be more precise than measurement B.

We can generalize for any number of measurements of a quantity x:

$$x_{wav} = \frac{\sum w_i x_i}{\sum w_i} \quad with: \quad w_i = \frac{1}{\sigma_i^2} \qquad (17)$$

Where: i = 1, 2, ……..,N

The weighted average is a function of the measured values x_i. Therefore, the uncertainty in the weighted average can be calculated using the error propagation expressions (see the document on Error Analysis in Experimental Physical Sciences)

➔ **Q2**. Prove that the uncertainty in x_{wav} is given by: $\sigma_{wav} = \frac{1}{\sqrt{\sum w_i}}$.

5. Investigating the mathematical relationship between two variables

In the vast majority of experiments you will perform this year, you'll be asked to determine one physical quantity y (the dependent variable) as a function of some other quantity x (the independent variable).

To accomplish this goal, you'll take a series of N measurements of the pair (x_i,y_i), where i = 1, 2,….., N. You will then find a function y = y(x) that describes the relation between the two measured quantities.

5.1. The linear regression method

If the two variables are clearly related by a linear relationship such as: y(x) = a + bx, we have to consider the *linear regression method* to determine the most probable values of parameters a and b. Linear regression is a *method of least squares.*

For any value x_i, if we knew coefficients a and b we could calculate the true value of the corresponding y_i

$$y_{i,true} = a + bx_i \tag{18}$$

Measured y_i values usually obey Gauss' distribution. The probability of obtaining y_i is given by:

$$P_{a,b}(y_i) \approx \frac{1}{\sigma_y} e^{-(y_i - a - bx_i)^2 / 2\sigma_y^2} \tag{19}$$

where σ_y is the width of the y-values distribution.

Probability of obtaining the complete set of measurements: $y_1, y_2, \ldots, y_N$ is given by the product:

$$P_{a,b}(y_1, y_2, ...y_N) = P_{a,b}(y_1)P_{a,b}(y_2)....P_{a,b}(y_N) \approx \frac{1}{\sigma_y^N} e^{-\frac{\chi^2}{2}} \tag{20}$$

Exponent is called *chi-squared*:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - a - bx_i)^2}{\sigma_y^2} \tag{21}$$

The *principle of maximum likelihood* requires a maximum probability or a minimum chi-squared. Setting χ^2 to minimum means:

$$\frac{\partial \chi^2}{\partial a} = 0 = -\frac{2}{\sigma_y^2} \sum_{i=1}^{N} (y_i - a - bx_i) \tag{22a}$$

$$\frac{\partial \chi^2}{\partial b} = 0 = -\frac{2}{\sigma_y^2} \sum_{i=1}^{N} x_i (y_i - a - bx_i) \tag{22b}$$

Equations (22a,b) can be solved for a and b:

$$a = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - \left(\sum x_i\right)^2} \tag{23a}$$

$$b = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - \left(\sum x_i\right)^2} \qquad (23b)$$

➔ **Q3.** Prove (23a,b)

Equations 23a and 23b are best estimates for coefficients a and b, based on N measurements (x_i, y_i) where i = 1, 2, …,N. The straight line y = a + bx is called *least-squares fit to the data* or *the line of regression of y on x.*

5.2. Uncertainties in measured values and calculated parameters

Measurements of y_i are normally distributed about the true value $y_{i,true} = a + bx_i$.
Distribution is characterized by the width parameter σ_y.
Therefore, the *deviations* $(y_i - y_{i,true}) = (y_i - a - bx_i)$ will also be normally distributed, but about zero. The distribution of deviations will have the same width σ_y as the measurements distribution. This suggests that $\sigma_y^2 = \frac{\sum (y_i - a - bx_i)^2}{N}$ (24)

The above result is not complete unless we determine the uncertainties in calculated coefficients a and b. To do this, we use the error propagation equation for parameters a and b, assuming $a = a(y_i), b = b(y_i)$ *and* $y_i = f(x_i)$:

$$\sigma_a^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial a}{\partial y_i}\right)^2 \qquad (25)$$

In (25), σ_i are standard deviations of data points y_i. We assumed no uncertainty in x_i.
Calculating $\frac{\partial a}{\partial y_i}$ *and* $\frac{\partial b}{\partial y_i}$ from (23a, b), we obtain:

$$\sigma_a^2 = \frac{1}{\Delta}\sum \frac{x_i^2}{\sigma_i^2}$$

$$\sigma_b^2 = \frac{1}{\Delta}\sum \frac{1}{\sigma_i^2} \qquad (26)$$

$$where: \quad \Delta = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2}\right)^2$$

➔ **Q4**. Prove (26)

When analyzing experimental data, we have to take into account a parameter called *number of degrees of freedom*, which is the number of data points (N) minus the number of parameters (m), calculated from the fit. This suggests that instead of (24), we can use a better estimate of σ_y for a linear fit:

$$\sigma_y^2 = \frac{1}{N-2}\sum (y_i - a - bx_i)^2 \qquad (27)$$

σ_y is also called variance. It has the property that $\sigma_y^2 = s^2$

5.3. The least squares fitting method applied to other functions

For problems in which the fitting function is *linear in the parameters*, the methods of least squares can be extended to any number of terms such as a power series polynomial:

$$y(x) = \sum_{k=1}^{m} a_k x^{k-1} \qquad (28)$$

The solution to this problem could be very tedious: 'm' coupled linear equations for the 'm' parameters have to be solved. a_1, a_2, ..., a_m are written in the form of m × m determinants.

The least squares method can work well with simple exponential functions, which are very important in physics:

$$y = Ae^{Bx} \qquad (29)$$

Examples:

- Intensity I of the radiation decays after traveling a distance x through a medium (shield) and is given by:

$$I = I_0 e^{-\mu x} \qquad (30a)$$

where I_0 is the original intensity and μ characterizes the attenuation by the medium.

- Charge on a capacitor in series with a resistor decays exponentially:

$$Q = Q_0 e^{-\frac{t}{\tau}} \qquad (30b)$$

where Q_0 is the original charge and $\tau = RC$ where R is resistance and C is capacitance.

To apply linearization to (29), we simply take the natural logarithm of both sides:

$$\ln y = \ln A + Bx$$

We can see that even if y is not linear in x, lny is.

With: z =lny, a = lnA and b = B, we can write:

$$z = a + bx$$

5.4. Non-linear fitting

Analytic methods of least-squares fitting used before for linear or lineralized functions cannot be applied to non-linear fitting problems. The probability function (Eq. 20) can be generalized to m parameters by the following approximation:

$$P(a_1, a_2, ..., a_m) \approx \exp\left\{-\frac{1}{2}\sum\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\} \qquad (31)$$

As before, we have to maximize the likelihood with respect to the parameters a_1, a_2, .., a_m, or minimize the exponent χ^2, also called *goodness-of-fit parameter*:

$$\chi^2 = \sum\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2 \qquad (32)$$

x_i and y_i are the measured variables, σ_i are the uncertainties in y_i and $y(x_i)$ are values of the calculated fit function at x_i which depends on the parameters a_1, a_2, .., a_m. The method of least squares states that the optimum values of the parameters a_1, a_2, .., a_m are calculated by minimizing χ^2 with respect to each parameter. This yields m coupled equations in the m parameters.
The coupled equations may not be linear in all the parameters. In this case, we must treat χ^2 as a continuous function of the m parameters and search the m-dimensional space for the minimum of χ^2. An alternative method would be to find the roots of the m nonlinear, coupled equations by using approximation methods. Both approaches are difficult. Several (very useful) computer routines will be introduced in the computational part of this course.

5.5. Covariance and Correlation: how two variables are related through their errors
Covariance is part of the 'leastsq' module output from scipy. Covariance is a parameter σ_{xy} which, if not zero, assesses that the errors in x and y are correlated.
The *coefficient of linear correlation* is a measure of the goodness of the fit.

Assume we measured x_i and y_i, N times. We calculated the mean values $\bar{x}, \bar{y}$ and the corresponding standard deviations σ_x and σ_y. We found out that uncertainties are small and deviations from the mean values are also small for both x and y.
Assume our experiment aims at finding a value for a function f(x,y) which takes values for different pairs x_i, y_i: $f_i = f(x_i, y_i) \quad i = 1,2,....,N$
Given small deviations from the mean, we can expand f(x,y) around $\bar{x}$ *and* $\bar{y}$:

$$f_i = f(x_i, y_i) \cong f(\bar{x}, \bar{y}) + \left.\frac{\partial f}{\partial x}\right|_{x=\bar{x}} (x_i - \bar{x}) + \left.\frac{\partial f}{\partial y}\right|_{y=\bar{y}} (y_i - \bar{y}) \qquad (33)$$

The average of all function values can be calculated as any other one-variable mean value:

$\bar{f} = \dfrac{\sum_{i=1}^{N} f_i}{N}$. Using (33), we notice that the partial derivatives are taken at $x_i = \bar{x}$ *and* $y_i = \bar{y}$ and are the same for all i-values. The second and third terms from (33) become zero upon summation over all i values and we are left with only:

$$\bar{f} = f(\bar{x}, \bar{y}) \qquad (34)$$

The mean of function f is the value calculated at $x = \bar{x}$ *and* $y = \bar{y}$.
It is not unreasonable to calculate the variance associated with the N values of function f:

$$\sigma_f^2 = \frac{\sum_{i=1}^{N} (f_i - \bar{f})^2}{N} \qquad (35)$$

Substituting (33) and (34) we obtain:

$$\sigma_f^2 = \frac{1}{N}\sum\left[\left(\frac{\partial f}{\partial x}\right)(x_i - \bar{x}) + \left(\frac{\partial f}{\partial y}\right)(y_i - \bar{y})\right]^2$$
$$= \left(\frac{\partial f}{\partial x}\right)^2 \frac{1}{N}\sum(x_i - \bar{x})^2 + \left(\frac{\partial f}{\partial y}\right)^2 \frac{1}{N}\sum(y_i - \bar{y})^2 + \frac{2}{N}\left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right)\sum(x_i - \bar{x})(y_i - \bar{y}) \quad (36)$$

The sums in the first and second terms from (36) define the standard deviations σ_x and σ_y. σ_{xy} is called *sample covariance of x and y:*

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \quad (37)$$

Covariance is a quantity different from zero only if variables x and y are dependent. If x and y are independent, (36) reduces to the usual relation for error propagation. When σ_{xy} is not zero, the errors in x and y are correlated. σ_{xy} can take positive or negative values but it can be proved that:

$|\sigma_{xy}| \le \sigma_x \sigma_y$ (Schwarz's inequality) (38)

The *linear correlation coefficient r* measures the extent to which the set (x_i, y_i) (i = 1,2,...N) supports a linear relation between x and y:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (39)$$

Values of r near +1 or -1 indicate strong correlation; values close to zero indicate very little or zero correlation.

5.6. The chi-squared test for goodness of fit

Chi-squared have been used before in relation to how well the observed values fit a certain function. Assume we measured N pairs (x_i, y_i) (i = 1,2,...N) and x_i had negligible uncertainty. $f(x_i)$ is a linear function in m parameters defined as the expected value of y_i. To test how well y fits the function f(x) we calculate:

$$\chi^2 = \sum_{i=1}^{N}\frac{[y_i - f(x_i)]^2}{\sigma_{y\,i}^2} \quad (40)$$

To test the agreement between a dependent variable and a function, we have to have a closer look at the constraints imposed by the calculation itself.

The *variance of the fit* is defined as:

$$s^2 = \frac{1}{N-m}\frac{\sum_{i=1}^{N}\frac{[y_i - f(x_i)]^2}{\sigma_{y_i}^2}}{\frac{\sum_{i=1}^{N}\frac{1}{\sigma_{y_i}^2}}{N}} \quad (41)$$

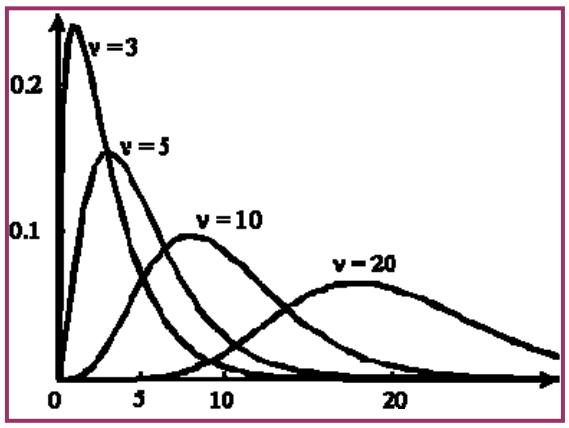
In (41), the factor ν = N-m is *the number of degrees of freedom* for fitting N data points with m parameters.
The relationship between the variance of the fit (41) and chi-squared (40) is given by:

$\frac{\chi^2}{\nu} = \frac{s^2}{\left\langle \sigma_i^2 \right\rangle}$, where $\left\langle \sigma_i^2 \right\rangle$ is the weighted average of all the individual variances:

$$\left\langle \sigma_i^2 \right\rangle = \left[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right]^{-1}$$

s^2 is characteristic for both the distribution of data and the goodness of the fit.
If the fitting function is a good approximation to experimental data, the estimated variance s^2 should agree well with σ^2 and the reduced chi-squared $\chi_\nu^2 = \frac{\chi^2}{\nu}$ should be close to 1.
A value of χ_ν^2 much larger than 1 or much less than 1 means an underestimation or overestimation of experimental uncertainties, respectively.

Another valuable evaluation tool for goodness of the fit comes from the probability distribution of chi-squared values. The probability distribution function for χ^2 with ν degrees of freedom is presented in many textbooks on statistics or error analysis (see Bevington 11.1). Below, you may see this distribution for several values of the number of degrees of freedom ν.



Let's say we did the same experiment 1000 times, each time we calculated the chi-squared value and plotted them all on a graph. The x-axis is the chi-squared value; the y-axis is the number of individual experiments that yielded that chi-square value
If the results were perfect we should have obtained a chi-square value of zero because the observed and calculated values were identical. This never happens in real lab experiments. Most times the fit is very different from experimental values. This is represented by the long tail on the graph.
The main properties of the chi-squared distribution are:

- The distribution is constructed so that the total area under the curve is equal to 1.
- The mean of the distribution is equal to the number of degrees of freedom.
- As the degrees of freedom increase, the chi-square curve approaches a normal distribution.

For a good fit the reduced chi-squared $\chi_\nu^2 = \frac{\chi^2}{\nu}$ should be close to 1.

6. The Poisson distribution or distribution of rare events

A typical example of a Poisson experiment is the statistical study of a radioactive decay: a radioactive source is placed in front of a counter (Geiger-Mueller tube) and random events are recorded. Every time ionizing radiation passes through the counter it produces ionization in the gas filling the tube and a negative charge is accumulated at its anode. The anode is connected through a resistor to the power supply and thus a negative voltage pulse appears at the anode and is counted as 'event'.

All Poisson experiments are characterized by the following properties:

1) The experiment results in "success" or "failure" outcomes. Success is determined by the physical recording of a radioactive event; failure is obviously the non-event.
2) If we investigate the decay within a certain time or energy window, we can define the average number of successes (μ) that occur per window.
3) It is clear that the probability of a success occurrence is proportional to the size of the window.
4) By reducing the size of the window to something very, very small, the success probability will be close to zero.

Letting μ be the mean number of successes that occur in a specified window and x be the actual number of successes in that window (x is also called a Poisson random variable), we can define the Poisson probability:

$$P(x;\mu) = \frac{e^{-\mu}(\mu^x)}{x!} \qquad (43)$$

This expression gives the probability distribution of the Poisson random variable x (for a complete derivation, see Bevington, Ch. 2, p.23-25). The Poisson distribution is discrete, defined only at integer values of the variable, in contrast with the continuous normal (Gaussian) distribution.

The variance of a Poisson distribution is equal to the mean μ. The standard deviation is given by $\sigma = \sqrt{\mu}$ which is the golden rule of assessing the uncertainty in a counting experiment: it is given by the root of the average number of counts per counting interval.

For large values of μ, we don't see "rare events" anymore and the probability sum:

$$S_P(x_1, x_2; \mu) = \sum_{x_1}^{x_2} \frac{e^{-\mu}(\mu^x)}{x!} \qquad (44)$$

may be approximated by an integral of the Gaussian function (see Bevington p. 24-25).

Some exercises

1) Assume you measured the period of a pendulum 12 times, under identical conditions and you have obtained the following data (in sec.):
   1.31, 1.35, 1.34, 1.37, 1.41, 1.37, 1.32, 1.33, 1.35, 1.36, 1.35, 1.33
   a) Find the mean and the standard deviation of the data set
   b) Calculate the uncertainty in the mean value.

2) Suppose we are interested in a quantity f(x, y, z) made up of three independent components that we measured as: x = 1.20 $\pm$ 0.42, y = 2.71 $\pm$ 0.01, z = 0.010 $\pm$ 0.001. The expression of f is given by: $f = \frac{x+y}{x+z}$. Express f $\pm$ Δf .

3) In a lab exercise, you measured the acceleration of a cart on a low-friction track slope by timing the passage of the cart through two photocells separated by a distance s. The cart has a length l, needs a time t_1 to pass through the first photocell and a time t_2 to pass through the second one. Given: l = 4.00 $\pm$ 0.05 cm, s = 110.0 $\pm$ 0.2 cm, t_1 = 0.055 $\pm$ 0.01 sec, t_2 = 0.033 $\pm$ 0.001 sec., calculate the acceleration and its uncertainty.

4) Imagine that you received a "black box device" with one input and one output. Upon applying an input voltage V_1, you measured an output V_2. Data are presented in the following table:

V_1 (V)	20	23	23	22
V_2 (mV)	30	32	35	31

   a) Calculate the variances σ_x^2 and σ_y^2 and the covariance σ_{xy} .
   b) Calculate the coefficient of linear correlation r and discuss the output of your device.

5) Given five data pairs:

   x = 1 2 3 4 5
   y = 4 6 3 0 2

   do the following:
   a) Draw a scatter plot and the least-squares line that fits the points
   b) Calculate the correlation coefficient and decide if data pairs are strongly correlated or not

6) Fiesta plates are known to be radioactive due to the presence of low amounts of uranium oxide in the glaze. In order to measure the activity of such a plate, a student counts 48 events in 5 minutes. The background was measured independently to be 16 counts in 2 minutes. Is there significant evidence that the plate is radioactive?

7) In doing a radioactive measurement of an air sample, a student knows that the expected number of counts per minute is 3. The enclosed table shows the number of decays (n) observed by the student in 100 separate intervals of 1 minute each:

Times observed	4	16	28	31	15	8	3	2	1	0
n	0	1	2	3	4	5	6	7	8	9

a) Assuming that the events are Poisson-distributed, calculate the probability that 0, 1, 2,... events are measured ($P_3(n)$).
b) Plot the fraction of times the result n was found (f_n) vs. n. On the same graph plot $P_3(n)$ vs. n. Does the expected distribution seem to fit the data?
c) Calculate χ^2 and examine the χ^2 probability for the whole data set. Discuss the result.