Basic Error Analysis

Physics 401
Fall 2018
Eugene V Colla
Agenda

• Errors and uncertainties
• The Reading Error
• Accuracy and precession
• Systematic and statistical errors
• Fitting errors
• Appendix. Working with oil drop data
  • Nonlinear fitting
What and when we need to know about errors. Everyday life.

$T = 63^\circ F \pm ?$  $\rightarrow$  Best guess $\Delta T \sim 0.5^\circ F$

Wind speed 4mph$\pm ?$  $\rightarrow$  Best guess $\pm 0.5mph$
What and when we need to know about errors. Industry.

Clearance fit

[Diagram showing tolerance zones, upper and lower deviations, and clearance fit examples]

Fastener
MMC = 0.747
LMC = 0.744

Hole
MMC = 0.750
LMC = 0.753
1675 Ole Roemer: 220,000 Km/sec

Measurement of the speed of the light

1644-1710

Does it make sense? What is missing?

NIST Bolder Colorado $c = 299,792,456.2 \pm 1.1 \text{ m/s.}$
We do not care about accuracy better than 1mm.

If ruler is not okay, we need to use digital caliper.

Probably the natural limit of accuracy can be due to length uncertainty because of temperature expansion. For 53mm $\Delta L \approx 0.012 \text{mm/K}$.

Reading Error $= \pm \frac{1}{2}$ (least count or minimum gradation).
Reading error. Digital meters.

Fluke 8845A multimeter

Example Vdc (reading) = 0.85V

$$\Delta V = 0.83 \times (1.8 \times 10^{-5})$$

$$+1.0 \times (0.7 \times 10^{-5}) \approx 2.2 \times 10^{-5}$$

$$= 22 \mu V$$

8846A Accuracy

Accuracy is given as ± (% measurement + % of range)

<table>
<thead>
<tr>
<th>Range</th>
<th>24 Hour (23 ±1 °C)</th>
<th>90 Days (23 ±5 °C)</th>
<th>1 Year (23 ±5 °C)</th>
<th>Temperature Coefficient/ °C Outside 18 to 28 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 mV</td>
<td>0.0025 + 0.003</td>
<td>0.0025 + 0.0035</td>
<td>0.0037 + 0.0035</td>
<td>0.0005 + 0.0005</td>
</tr>
<tr>
<td>1 V</td>
<td>0.0018 + 0.0006</td>
<td>0.0018 + 0.0007</td>
<td>0.0025 + 0.0007</td>
<td>0.0005 + 0.0001</td>
</tr>
<tr>
<td>10 V</td>
<td>0.0013 + 0.0004</td>
<td>0.0018 + 0.0005</td>
<td>0.0024 + 0.0005</td>
<td>0.0005 + 0.0001</td>
</tr>
<tr>
<td>100 V</td>
<td>0.0018 + 0.0006</td>
<td>0.0027 + 0.0006</td>
<td>0.0038 + 0.0006</td>
<td>0.0005 + 0.0001</td>
</tr>
<tr>
<td>1000 V</td>
<td>0.0018 + 0.0006</td>
<td>0.0031 + 0.001</td>
<td>0.0041 + 0.001</td>
<td>0.0005 + 0.0001</td>
</tr>
</tbody>
</table>
The accuracy of an experiment is a measure of how close the result of the experiment comes to the true value.

Precision refers to how closely individual measurements agree with each other.
Accuracy and precession

- Not Precise, Not Accurate
- Precise, Not Accurate
- Accurate, Not Precise
- Accurate, Precise
- Errors
**Systematic and random errors**

- **Systematic Error**: reproducible inaccuracy introduced by faulty equipment, calibration or technique.

- **Random errors**: Indefiniteness of results due to finite precision of experiment. Measure of fluctuation in result after repeatable experimentation.

Sources of systematic errors: poor calibration of the equipment, changes of environmental conditions, imperfect method of observation, drift and some offset in readings etc.

Example #1: measuring of the DC voltage

\[ U = R \times I \]

In expectation, the actual result depends on:

\[ E_{\text{off}} = f(\text{time}, \text{temperature}) \]

\[ U = \frac{R \times I - \left( \frac{R}{R_{\text{in}}} \right) E_{\text{off}}}{1 + \left( \frac{R}{R_{\text{in}}} \right)} \]

where:

- \( U \) is the measured voltage
- \( R \) is the resistance
- \( I \) is the current
- \( E_{\text{off}} \) is the offset
- \( R_{\text{in}} \) is the input resistance

\( E_{\text{off}} \) is a function of time and temperature.
Example #3: poor calibration

Measuring of the speed of the second sound in superfluid He4

Published data
$T_\lambda = 2.17\, \text{K}$

P403 results
$T_\lambda = 2.1\, \text{K}$

$U_2$ (m/s)

Temperature sensor

Resonator

LHe

10μA

HP34401A
DMM

DT-470/471-50

LakeShore
Random errors

Result of measurement

\[ X_{\text{meas}} = X_{\text{true}} + e_s + e_r \]

Correct value

Systematic error

Random error

\[ e_s = 0 \]

physics 401
A statistical process is described through a Poisson Distribution if:

- **random process** → for a given nucleus probability for a decay to occur is the same in each time interval.
- **universal probability** → the probability to decay in a given time interval is same for all nuclei.
- **no correlation between two instances** (the decay of one nucleus does not change the probability for a second nucleus to decay.)

**Poisson Distribution**

\[ P_n(t) = \frac{(rt)^n}{n!} e^{-rt} \quad n = 0,1,2,... \] 

**Variables**

- **r**: decay rate [counts/s]
- **t**: time interval [s]
- **P_n(rt)**: Probability to have **n** decays in time interval **t**

**Graph**

- **P vs. number of counts**
  - **rt=1**
  - **rt=4**
  - **rt=10**
Poisson distribution

\[ P_n(t) = \frac{(rt)^n}{n!} e^{-rt} \quad n = 0, 1, 2, \ldots \]

- \( r \): decay rate [counts/s]
- \( t \): time interval [s]

\( P_n(rt) \): Probability to have \( n \) decays in time interval \( t \)

**Properties of the Poisson distribution:**

\[ \sum_{n=0}^{\infty} P_n(rt) = 1 \quad \text{probabilities sum to 1} \]

\[ < n > = \sum_{n=0}^{\infty} n \cdot P_n(rt) = rt \quad \text{the mean} \]

\[ \sigma = \sqrt{\sum_{n=0}^{\infty} (n - < n >)^2 P_n(rt)} = \sqrt{rt} \quad \text{standard deviation} \]
Poisson distribution at large $rt$

$$P_n(t) = \frac{(rt)^n}{n!} e^{-rt} \quad n = 0, 1, 2, ...$$

Poisson and Gaussian distributions

- "Poisson distribution"
- "Gaussian distribution"

Gaussian distribution:
continuous

$$P_n(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$
Normal (Gaussian) distribution

Error in the mean is given as $\frac{\sigma}{\sqrt{N}}$

$N$ – number of events
Measurement in presence of noise

Source of noisy signal

Actual measured values:
- 4.89855
- 5.25111
- 2.93382
- 4.31753
- 4.67903
- 3.52626
- 4.12001
- 2.93411

Expected value 5V
Measurement in presence of noise

10

$\mu = 5.2924 \text{V}$

100

$\mu = 5.139 \text{V}$

$10^4$

$\mu = 4.992 \text{V}$

$10^6$

$\mu = 5.003 \text{V}$
**Result**

\[ U = x_c \pm \frac{\sigma}{\sqrt{N}} \]

\( \sigma \) - standard deviation

\( N \) – number of samples

For \( N=10^6 \) \( U=4.999\pm0.001 \)  

0.02% accuracy
Ag $\beta$ decay

$^{108}\text{Ag}$ $t_{1/2}=157\text{s}$

$^{110}\text{Ag}$ $t_{1/2}=24.6\text{s}$

Model ExpDec2
Equation $y = A_1 \exp(-x/t_1) + A_2 \exp(-x/t_2) + y_0$
Reduced Chi-Sqr 1.43698
Adj. R-Square 0.96716

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_0$</td>
<td>0.02351</td>
<td>0.95435</td>
</tr>
<tr>
<td>$A_1$</td>
<td>104.87306</td>
<td>12.77612</td>
</tr>
<tr>
<td>$t_1$</td>
<td>177.75903</td>
<td>18.44979</td>
</tr>
<tr>
<td>$A_2$</td>
<td>710.01478</td>
<td>25.44606</td>
</tr>
<tr>
<td>$t_2$</td>
<td>30.32479</td>
<td>1.6525</td>
</tr>
</tbody>
</table>

Residuals

Model Gauss
Equation $y = y_0 + \frac{A}{(w \sqrt{\pi/2})} \exp\left(-2\left(\frac{x-x_c}{w}\right)^2\right)$
Reduced Chi-Sqr 4.77061
Adj. R-Square 0.93464

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_0$</td>
<td>1.44204</td>
<td>0.48702</td>
</tr>
<tr>
<td>$x_c$</td>
<td>1.49992</td>
<td>0.19171</td>
</tr>
<tr>
<td>$w$</td>
<td>5.93398</td>
<td>0.40771</td>
</tr>
<tr>
<td>$A$</td>
<td>219.24559</td>
<td>14.47587</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2.96699</td>
<td></td>
</tr>
<tr>
<td>$\text{FWHM}$</td>
<td>6.98673</td>
<td></td>
</tr>
<tr>
<td>$\text{Height}$</td>
<td>29.4798</td>
<td></td>
</tr>
</tbody>
</table>

$y = A_1 \exp\left(\frac{-t}{t_1}\right) + A_2 \exp\left(\frac{-t}{t_2}\right) + y_0$
Fitting. Analysis of the residuals

Ag β decay

Test 1. Fourier analysis

No pronounced frequencies found
Fitting. Analysis of the residuals

Ag β decay

Correlation function

\[ y(m) = \sum_{n=0}^{M-1} f(n)g(n-m) \]

autocorrelation function

\[ y(m) = \sum_{n=0}^{M-1} f(n)f(n-m) \]
Fitting. Analysis of the residuals. Non “ideal” case

Ag β decay

<table>
<thead>
<tr>
<th>Model</th>
<th>ExpDec2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation</td>
<td>y = A1 exp(-x/t1) + A2 exp(-x/t2) + y0</td>
</tr>
<tr>
<td>Reduced Chi-Sqr</td>
<td>100.10041</td>
</tr>
<tr>
<td>Adj. R-Square</td>
<td>0.99181</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>y0</td>
<td>5.18284</td>
<td>1.96542</td>
</tr>
<tr>
<td>A1</td>
<td>130.85555</td>
<td>20.27379</td>
</tr>
<tr>
<td>t1</td>
<td>145.8949</td>
<td>21.82649</td>
</tr>
<tr>
<td>A2</td>
<td>792.62197</td>
<td>19.21953</td>
</tr>
<tr>
<td>t2</td>
<td>27.83939</td>
<td>1.30097</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>time (s)</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>800</td>
</tr>
<tr>
<td>100</td>
<td>600</td>
</tr>
<tr>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>400</td>
<td>0</td>
</tr>
</tbody>
</table>

| t1(s) | 177.76 |
| t2(s) | 30.32 |

Data + “noise”:

| t1(s) | 145.89 |
| t2(s) | 27.94 |
Ag $\beta$ decay

Histogram does not follow the normal distribution and there is frequency of 0.333 is present in spectrum
Conclusion: fitting function should be modified by adding an additional term:

\[ y(t) = y_0 + A_1 \exp\left(\frac{-t}{t_1}\right) + A_2 \exp\left(\frac{-t}{t_2}\right) + A_3 \sin(\omega t + \theta) \]
Fitting. Analysis of the residuals. Non “ideal” case

**Clear experiment** | **Data + noise** | **Modified fitting**
---|---|---
$t_1(s)$ | 177.76 | 145.89 | 172.79
$t_2(s)$ | 30.32 | 27.94 | 30.17
In general we could expect both components of errors

\[ Q_{\text{meas}} = Q_{\text{true}} + e_s + e_r \]

\( e_s \) - systematic error comes from uncertainties of plates separation distance, applied DC voltage, ambient temperature etc.

\[ V = V_{DC} \pm \Delta V, \quad d = d_0 \pm \Delta d \quad \ldots \]

\( e_r \) - random errors are related to uncertainty of the knowledge of the actual \( t_g \) and \( t_{\text{rise}} \).

Uncertainty of time of crossing the marker line. It is random.
Systematic component. Error propagation.
Millikan oil drop experiment.

\[ X_{\text{meas}} = X_{\text{true}} + e_s + e_r \quad \text{Systematic error} \]

\[ Q = F \cdot S \cdot T = \left( \frac{1}{f_c^{3/2}} \right) \frac{9 \pi d}{V} \sqrt{\frac{2\eta^3 x^3}{g \rho}} \frac{1}{\sqrt{t_g}} \left( \frac{1}{t_g} + \frac{1}{t_{\text{rise}}} \right) \]

\[ F = \frac{1}{f_c^{3/2}} \approx 1 - \left( \frac{t_g}{\tau_g} \right)^2 \]

\[ S = \frac{9 \pi d}{V} \sqrt{\frac{2\eta^3 x^3}{g \rho}} \]

\[ T = \frac{1}{\sqrt{t_g}} \left( \frac{1}{t_g} + \frac{1}{t_{\text{rise}}} \right) \]

\[ \Delta Q = \sqrt{\left( \frac{dQ}{dF} \right)^2 (\Delta F)^2 + \left( \frac{dQ}{dS} \right)^2 (\Delta S)^2 + \left( \frac{dQ}{dT} \right)^2 (\Delta T)^2} \approx \sqrt{\left( \frac{dQ}{dS} \right)^2 (\Delta S)^2 + \left( \frac{dQ}{dT} \right)^2 (\Delta T)^2} \]

\[ = \sqrt{(FT)^2 (\Delta S)^2 + (FS)^2 (\Delta T)^2} = Q \sqrt{\left( \frac{\Delta S}{S} \right)^2 + \left( \frac{\Delta T}{T} \right)^2} \]

$$X_{\text{meas}} = X_{\text{true}} + e_s + e_r$$

$$\Delta Q \approx Q \sqrt{\left( \frac{\Delta S}{S} \right)^2 + \left( \frac{\Delta T}{T} \right)^2}$$

$$\frac{\Delta S}{S} = \sqrt{\left( \frac{\Delta d}{d} \right)^2 + \left( \frac{\Delta V}{V} \right)^2 + \left( \frac{3}{2} \frac{\Delta x}{x} \right)^2 + \left( \frac{3}{2} \frac{\Delta \eta}{\eta} \right)^2 + \left( \frac{1}{2} \frac{\Delta \rho}{\rho} \right)^2 + \left( \frac{1}{2} \frac{\Delta g}{g} \right)^2} \approx \sqrt{\left( \frac{\Delta d}{d} \right)^2 + \left( \frac{3}{2} \frac{\Delta x}{x} \right)^2}$$

$$\Delta T = \sqrt{\left( \frac{3/2}{t_g^{5/2}} + \frac{1/2}{t_g^{3/2}} \frac{1}{t_{\text{rise}}} \right)^2 \Delta t_g^2 + \left( \frac{1}{t_g^{1/2}} \frac{1}{t_{\text{rise}}} \right)^2 \Delta t_{\text{rise}}^2}$$
Appendix #1. Analyzing of the statistical data.

Step 1. Origin Project For Raw Data:
\\engr-file-03\PHYINST\APL Courses\PHYCS401\Students\2. Millikan Raw Data

Here should only the files with raw data but not other files which you using for calculations. All other files you can save in your personal folder.

All these projects with raw data should be stored in:
\\engr-file-03\PHYINST\APL Courses\PHYCS401\Students\2. Millikan Raw Data
### Appendix #1. Analyzing of the statistical data.

#### Step 2. Working on personal Origin project

Make a copy of the Millikan1 project to your personal folder and open it.

Paste these 5 parameters and raw data from Section L1-L4.opj projects.

- Calculate manually the actual air viscosity.
- Prepare equations calculations of data in next columns (Set column values...), Switch Recalculate in Auto mode.

```
<table>
<thead>
<tr>
<th>Long Name</th>
<th>Units</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parameter label</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Par</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_g</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>t_r</td>
<td>s</td>
<td></td>
</tr>
<tr>
<td>r_c</td>
<td>m</td>
<td></td>
</tr>
<tr>
<td>tau_g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Example equations:
- \( r_c = \frac{6.18 \times 10^{-8}}{\rho [mmHg]} \)
- \( r_g = \frac{2\eta x}{\rho gr_c^2} \)
- \( F = \frac{1}{r_c^{3/2}} \approx 1 \)
Appendix #1. Analyzing of the statistical data.

Step 3. Histogram graph

First use the data from the column with drop charges and plot the histogram.
Step 4. Histogram. Bin size

Origin will automatically but not optimally adjust the bin size $h$. In this page figure $h=0.5$. There are several theoretical approaches how to find the optimal bin size.

$$h = \frac{3.5\sigma}{n^{1/3}}$$

$\sigma$ is the sample standard deviation and $n$ is total number of observation. For presented in Fig.1 results good value of $h \approx 0.1$
Step 4. Histogram. Bin size

To change the bin size click on graph and unplug the "Automatic Binning" option.

Bin size in this histogram is 0.1
Appendix #1. Analyzing of the statistical data.

Step 4. Find the bin Worksheet

Millikan oil drop experiment

Right click on histogram and choose “Go to Bin Worksheet”.

\[ n = \frac{Q}{1.602 \times 10^{-19}} \]
Appendix #1. Analyzing of the statistical data.

Step 5. Add Counts vs bin plot
Appendix #1. Analyzing of the statistical data.

Step 5. Multipeak Gaussian fitting

This plot can be used for peak fitting.
Step 5. Multipeak Gaussian Fitting

For $1^{st}$ peak $x_c \approx 0.882 \pm 0.007$
\( (x_i, y_i) \) is an experimental data array. \( x_i \) is an independent variable and \( y_i \) - dependent

\( f(x, \beta) \) is a model function and \( \beta \) is the vector of fitting (adjustable) parameters

The goal of the fitting procedure is to find the set of parameters which will generate the function \( f \) closest to the experimental points.

To reach this goal we will try to minimize the sum of squared deviation function \( (S) \):

\[
S(\beta) = \sum_{i=1}^{m} \left[ y_i - f(x_i, \beta) \right]^2
\]
Appendix #2. Fitting. The Choice of Parameters.

The goal of fitting is not only to find the curve best matching the experimental data but also to find the corresponding parameters which in majority cases are the important physical parameters.

There are several known mathematical algorithms for optimizing these parameters but in general the fitting procedure could have not only unique solution and the choice of initial parameters is very important issue.

\[ S(\beta) = \sum_{i=1}^{m} \left[ y_i - f(x_i, \beta) \right]^2 \]
Let we have the $S$ function dependent on parameter $\beta_i$ as shown on this graph.