Basic Error Analysis

Physics 401
Spring 2017
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.

\[ T = 63^\circ F \pm ? \quad \text{Best guess } \Delta T \sim 0.5^\circ F \]

Wind speed 4mph ±? \quad \text{Best guess } \pm 0.5mph
What and when we need to know about errors. Industry.

Clearance fit

Fastener
MMC = .747
LMC = .744

Hole
MMC = .750
LMC = .753

(a) Unilateral tolerance
(b) Bilateral tolerance.
Measurement of the speed of the light

1675 Ole Roemer: 220,000 Km/sec

Does it make sense? What is missing?

NIST Bolder Colorado c = 299,792,456.2±1.1 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).
**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

\[ U = R \times I \]

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

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

$T_\lambda = 2.1 \text{K}$

Temperature sensor

Resonator

LHe

10μA

HP34401A
DMM

Physics 401
Random errors

Result of measurement

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

Correct value

Systematic error

Random error

\( e_s = 0 \)

\( X_{\text{true}} \)

\( X_{\text{true}} \)

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.)

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

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

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

\[ \langle n \rangle = \sum_{n=0}^{\infty} n \cdot P_n(rt) = rt \], the mean

\[ \sigma = \sqrt{\sum_{n=0}^{\infty} (n - \langle n \rangle)^2 P_n(rt)} = \sqrt{rt} \], standard deviation
Poisson distribution at large $rt$

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

Poisson and Gaussian distributions

- "Poisson distribution"
- "Gaussian distribution"

Poisson distribution:
- Discrete
- Carl Friedrich Gauss (1777–1855)

Gaussian distribution: continuous
Normal (Gaussian) distribution

\[ P_n(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \]

Error in the mean is given as \[ \frac{\sigma}{\sqrt{N}} \]
Measurement in presence of noise

Source of noisy signal

Expected value 5V

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

Expected value 5V

Actual measured values
Measurement in presence of noise

Graph 1: $10^{10}$

Graph 2: $10^{100}$

Graph 3: $10^{4}$

Graph 4: $10^{6}$
Measurement in presence of noise

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

\( \sigma \) - standard deviation
\( N \) – number of samples

For \( N=10^6 \) \( U=4.999\pm0.001 \) 0.02% accuracy
Ag β decay

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

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

Model: ExpDec2
Equation:

$$y = A_1 \cdot \exp\left(-\frac{t}{t_1}\right) + A_2 \cdot \exp\left(-\frac{t}{t_2}\right) + y_0$$

Reduced Chi-Sqr: 1.43698
Adj. R-Square: 0.96716

<table>
<thead>
<tr>
<th>Parameter</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

Equation:

$$y = y_0 + \left(\frac{A}{w \sqrt{\pi / 2}}\right) \cdot \exp\left(-2 \cdot \left(\frac{x-x_c}{w}\right)^2\right)$$

Reduced Chi-Sqr: 4.77021
Adj. R-Square: 0.93464

<table>
<thead>
<tr>
<th>Parameter</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>FWHM</td>
<td>6.98673</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>29.4798</td>
<td></td>
</tr>
</tbody>
</table>
**Fitting. Analysis of the residuals**

**Ag β decay**

---

**Test 1. Fourier analysis**

No pronounced frequencies found

---

<table>
<thead>
<tr>
<th>Parameter</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>FWHM</td>
<td>6.98673</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>29.4798</td>
<td></td>
</tr>
</tbody>
</table>

---

The Fourier analysis did not reveal any pronounced frequencies.
Fitting. Analysis of the residuals

Ag $\beta$ decay

Test 1. Autocorrelation function

Correlation function

$$y(m) = \sum_{n=0}^{\infty} f(n)g(n-m)$$

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**


data table:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Clear experiment</th>
<th>Data + “noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1(s)$</td>
<td>177.76</td>
<td>145.89</td>
</tr>
<tr>
<td>$t_2(s)$</td>
<td>30.32</td>
<td>27.94</td>
</tr>
</tbody>
</table>
Ag β 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

<table>
<thead>
<tr>
<th></th>
<th>Clear experiment</th>
<th>Data + noise</th>
<th>Modified fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1(s)$</td>
<td>177.76</td>
<td>145.89</td>
<td>172.79</td>
</tr>
<tr>
<td>$t_2(s)$</td>
<td>30.32</td>
<td>27.94</td>
<td>30.17</td>
</tr>
</tbody>
</table>
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.

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

- \( e_r \) - 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.

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

$$\Delta Q = \sqrt{(S \bullet T)^2 \Delta F^2 + (F \bullet T)^2 \Delta S^2 + (F \bullet S)^2 \Delta T^2}$$

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

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

Step 1. Collect your data + parameters of the experiment in:
\\engr-file-03\\PHYINST\\APL Courses\\PHYCS401\\Students\\1. Millikan Oil Drop experiment\\Section L1m.opj

Use different columns for each student or team. This Origin project is for data collecting only but not for data analysis. For data analysis you have to copy these data and experiment parameters obtained by different students/team and paste it in one in your personal Origin project.

<table>
<thead>
<tr>
<th>Long Name</th>
<th>A(L)</th>
<th>B(Y)</th>
<th>C(Y)</th>
<th>D(Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units</td>
<td>Par</td>
<td>tg</td>
<td>tr</td>
<td></td>
</tr>
<tr>
<td>Comments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>765</td>
<td>15.56521</td>
<td>16.7815</td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>0.00145</td>
<td>23.07825</td>
<td>31.8955</td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>0.00317</td>
<td>20.14243</td>
<td>11.70129</td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>500</td>
<td>26.97377</td>
<td>22.47531</td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>20</td>
<td>16.34362</td>
<td>16.44208</td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>25.93429</td>
<td>25.02886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>15.34338</td>
<td>9.27446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>29.3815</td>
<td>19.6161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>student1, student2</td>
<td>26.0786</td>
<td>24.3434</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Setup and environmental parameters

Raw data

physics 401
Step 1. Slightly Modified Origin Project For Data Collection:

\texttt{\textbackslash engr\textbackslash file\textbackslash 03\textbackslash PHYINST\textbackslash APL Courses\textbackslash PHYCS401\textbackslash Students\textbackslash 1. Millikan Oil Drop experiment\textbackslash Section L1m.opj}

<table>
<thead>
<tr>
<th>A(L)</th>
<th>B(Y)</th>
<th>C(Y)</th>
<th>D(Y)</th>
<th>Y(Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter label</td>
<td>Par</td>
<td>tg</td>
<td>tr</td>
<td>n=Q/1.602e-19</td>
</tr>
<tr>
<td>student1, student2</td>
<td>student1, student2</td>
<td>student1, student2</td>
<td>student1, student2</td>
<td>student1 student2</td>
</tr>
</tbody>
</table>

Extra column for number of elementary charges
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.

\[ r_c[m] = \frac{6.18 \times 10^{-2}}{\rho [\text{mmHg}]} \]
\[ r_g = \frac{2 \eta x}{\rho g r_c^2} \]
\[ F = \frac{1}{r_c^{\frac{1}{2}}} \approx 1 \]
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 $h \sim 0.1$
Appendix #1. Analyzing of the statistical data.

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. Multipeak Gaussian fitting

Millikan oil drop experiment

To do this you have to add an extra plot to the graph Counts vs. Bin Center
Appendix #1. Oil Drop Data Issue.

Be careful with data selection obtained by different teams!
Appendix #1. oil Drop Data Issue.

Write-up, page 7. mistype in some copies

<table>
<thead>
<tr>
<th>quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viscosity of air</td>
<td>$\eta = 1.8478 \times 10^{-5}$ kg/m·s (25 ºC)</td>
</tr>
<tr>
<td>Density of oil</td>
<td>$\rho_{\text{oil}} = 886$ kg/m$^3$</td>
</tr>
<tr>
<td>Density of air</td>
<td>$\rho_{\text{air}} = 1.29$ kg/m$^3$</td>
</tr>
<tr>
<td>Acceleration due to gravity</td>
<td>$g = 9.801$ m/s$^3$</td>
</tr>
</tbody>
</table>

$\frac{\Delta \eta}{\Delta T} = 4.8 \times 10^{-8}$ kg / m·s / ºC

$\eta(T) = \eta(25^\circ C) + 4.8 \times 10^{-8} \times (T - 25)$ (kg/ m·s/ ºC)
Appendix #1. Analyzing of the statistical data.

Step 4. Multipeak Gaussian fitting

This plot can be used for peak fitting.

Millikan oil drop experiment
Step 4. Multipeak Gaussian fitting

This plot can be used for peak fitting.

Final result for first two peaks:
\( Q/e = 0.93 \pm 0.01 \)
\( Q/e = 1.87 \pm 0.07 \)

This pretty close to e and 2e

Here \( w = 2\sigma \) and error of the mean = \( \frac{\sigma}{\sqrt{N}} \approx \frac{\sigma}{\sqrt{A}} \)

Millikan oil drop experiment

\[ \text{Physics 401} \]
\[(x_i, y_i)\] 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 #1. 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.

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