Week 7

Combining Measurements

Goodness of Fits: some examples
Best way to Combine Numbers

• Suppose we have N measurements of the same physical quantity. How best to combine?

• Must test agreement or internal consistency

• Plot the data to get a visual sense for the numbers!

• cf. Bevington p. 56 ff.

Actual Muon Lifetime data from MPL:

<table>
<thead>
<tr>
<th>Trial</th>
<th>Lifetime $\mu$sec</th>
<th>Error $\mu$sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1138</td>
<td>0.0056</td>
</tr>
<tr>
<td>2</td>
<td>2.1245</td>
<td>0.0053</td>
</tr>
<tr>
<td>3</td>
<td>2.0916</td>
<td>0.0062</td>
</tr>
<tr>
<td>4</td>
<td>2.0871</td>
<td>0.0010</td>
</tr>
<tr>
<td>5</td>
<td>2.1040</td>
<td>0.0050</td>
</tr>
<tr>
<td>6</td>
<td>2.1140</td>
<td>0.0050</td>
</tr>
<tr>
<td>7</td>
<td>2.1136</td>
<td>0.0049</td>
</tr>
<tr>
<td>8</td>
<td>2.1000</td>
<td>0.0060</td>
</tr>
<tr>
<td>9</td>
<td>2.0990</td>
<td>0.0040</td>
</tr>
<tr>
<td>10</td>
<td>2.1000</td>
<td>0.0140</td>
</tr>
</tbody>
</table>
These numbers are not consistent... so what to do?

Cosmic Muon Lifetime (Both Charges Included)

- Straight average: 2.105 ± 0.006
- Average error: 2.105 ± 0.011
- Weighted mean: 2.102 ± 0.002
These numbers are not consistent... so what to do?

Cosmic Muon Lifetime (Both Charges Included)

- **Straight average**: Average error $2.105 \pm 0.006$
- **Standard average**: Standard deviation $2.105 \pm 0.011$

**Best compromise**: Weighted mean value - uses most information

- **Standard deviation uncertainty**: Honest evaluation of 'spread' of values
- **Error on mean**: $2.102 \pm 0.002$
Umm... weighted mean?

\[
P(\tau) = \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2} \left( \frac{\tau_i - \bar{\tau}}{\sigma_i} \right)^2} \right\} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2} \sum_{i=1}^{N} \left( \frac{\tau_i - \bar{\tau}}{\sigma_i} \right)^2}
\]

Maximum likelihood function: maximizing \( P \) and maximizing \( \ln P \) amount to the same thing

Assume each data point has a Gaussian error and that all measurements are independent

\[
\frac{d}{d\bar{\tau}} \left( -\frac{1}{2} \sum_{i=1}^{N} \left( \frac{\tau_i - \bar{\tau}}{\sigma_i} \right)^2 \right) = -\sum_{i=1}^{N} \frac{\tau_i - \bar{\tau}}{\sigma_i^2} = 0 \quad \text{Solve for best value of} \ \tau
\]

Weighted Mean

\[
\bar{\tau} = \frac{\sum_{i=1}^{N} \frac{\tau_i}{\sigma_i^2}}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}}
\]

two separate sums
Weighted Mean

\[
\bar{\tau} = \frac{\sum_{i=1}^{N} \frac{\tau_i}{\sigma_i^2}}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}}
\]

\[
\sigma_{\bar{\tau}}^2 = \sum_{i=1}^{N} \sigma_i^2 \left( \frac{\partial \bar{\tau}}{\partial \tau_i} \right)^2 + \text{correlations}
\]

How much does the answer change if the \(i\)th data point changes by a little bit?

Uncertainty on the Weighted Mean mnemonic formula

\[
\frac{1}{\sigma_{\bar{\tau}}^2} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}
\]
Why is my $\chi^2$ so bad? Case 1

$Y = FP1 + FP2 \, X$

$\Delta_i > \sigma_i$ leads to $\chi^2/\nu > 1$

Fit may be OK but the errors are underestimated

You may not inflate the $\sigma_i$'s simply to get to $\chi^2/\nu \sim 1$
Why is my $\chi^2$ so bad? Case 2

Why is my $\chi^2$ so bad? Case 2

\[ Y = FP1 + FP2 \times \]

Errors may be estimated correctly but the parent distribution from which data are drawn is not correct. The $\chi^2$ statistic "can't tell" if the data is bad or the fit function is bad.

\[ Y = FP1 + FP2 x + FP3 x^2 + FP4 x^3 \]

Adding terms will definitely reduce $\chi^2/\nu$, but caution: the fit values may be meaningless.
Why is my $\chi^2$ so bad? Case 2

$$y_{\text{fit}} = a_1 + a_2 x + a_3 x^2 + a_4 x^3$$

VALID NEW TERMS?

Is the change in $\chi^2$ enough better to justify extra terms but not “too good to be true”?

$$F_x = \frac{\chi^2(n = 2) - \chi^2(n = 3)}{\chi_v^2(n = 3)}$$

The “F-test” statistic see Bevington p.207

Ask: is $\sigma_{a_3} > |a_3|$? If so, the added term may not be meaningful.

Also, if $n \to N$, we can get $\chi^2 \to 0$, which is nonsense.
Why is my $\chi^2$ so bad? Case 3

$Y = FP1 + FP2 \times X + FP3 \times X^2$

$\chi^2 = \sum \left( \frac{Y - \text{Theory}}{\text{Error}} \right)^2$

<table>
<thead>
<tr>
<th>Entries</th>
<th>Constraints</th>
<th>d.o.f.</th>
<th>$\sigma(\chi^2 / \text{d.o.f.})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>3</td>
<td>17</td>
<td>0.34</td>
</tr>
</tbody>
</table>

$\Delta_i < \sigma_i$ leads to $\chi^2 / \nu < 1$

Fit is “no better” than a fit with $\chi^2 / \nu \sim 1$

$\sigma_i$’s have been overestimated; the data is better than you think
Why is my $\chi^2$ so bad? Case 4

Maybe you are just unlucky.

How close to "1" does $\chi^2_\nu$ have to be?

Without proof we give:

(\text{Ambramowicz \& Stegun p.943})

Thus, for a "good" fit:

\[
\sigma^2 = \frac{\sqrt{2}}{\nu}
\]

\[
\chi^2_\nu = 1 \pm \frac{\sqrt{2}}{\nu}
\]

e.g. for $N = 200$ $(x_i, y_i, \sigma_i)$ points fitted to a straight line $n=2$, then is "good" $\chi^2_\nu = 1 \pm \sqrt{2 / 198} = 1.00 \pm 0.10$
Why is my $\chi^2$ so bad? Case 5

• Maybe no amount of fiddling with functions and error bars will help... SYSTEMATIC may dominate your data
  - Non-linearities in devices, drifts in time, other untraceable irreproducibilities

• Remedy: Use you human judgement for how far you can shift parameter values and still have a reasonable “fit” that defines the uncertainty.
Combine numbers using the weighted mean, but be cautious about assigning a fair uncertainty.

Memorize the weighted mean formulas.

The $\chi^2$ statistic can’t distinguish between your messing up the errors or picking a bad parent function: You have got to figure it out.

$\chi^2$ is itself a statistical quantity, sometimes even a “good” fit has a bad $\chi^2$ (but don’t bet your career on it).
Backup Slides...