A Discussion of Errors and Uncertainties

R.L. Cooper
Physics 460: Modern Optics
Spring 2014
Experimental Errors

- The avoidable difference between “true” value and “measured” value
- Come in 2 types: Random and Systematic
- Example: Measure Resistance from I-V curve vs. Temperature
- Does this example seem reasonable?
Systematic Errors

• Typically experimental problems that “shift” the measurement away from its true value
• Example: Offset in voltmeter
  Example: Poor calibration
• No recipe to detect arbitrary “systematics”
• Rely upon experience, creativity, intuition, and a little luck to find and deal with them
• “Additional” measurements don’t help
Random Errors

- Fluctuations within multiple measurements
- Random in sign $x_{\text{true}} = x_i \pm \delta x_i$
- Multiple measurements will tend to average out these random errors
Question

• You make many measurements with only random errors. How many of your measurements $x_i$ are within the true value?

Assume they are normally distributed.
Answer

• About $\frac{1}{3}$ because $1\sigma$ is the uncertainty.
Describing a “Population”

- Suppose we don’t know the average or error
- Multiple measurements can be average

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

- Standard deviation estimates \( \sigma \)

\[
\sigma^2 = \frac{1}{N - 1} \sum_{i=1}^{N} (x_i - \bar{x})^2
\]
Describing a “Population” (cont.)

• How well do we know the average value?
• If \( \sigma \) is small, distribution is narrow \( \rightarrow \) most values will be “close” to average
• The more measurements we have, the more confident in distribution we are

\[
d\bar{x} = \frac{\sigma}{\sqrt{N}}
\]

• Weighted average is more sophisticated
Propagation of Errors

• Example: Resistance of simple resistor \( \Rightarrow \) measure voltage drop \( V \) and current through resistor \( I \). \[ R = \frac{V}{I} \]

• Both of these measurements will have some uncertainty

• We assume both uncertainties are independent of each other
Propagation of Errors

• Very similar to $\varepsilon-\delta$ in calculus $f(x, y)$

$$df = \sqrt{\left(\frac{df}{dx}\right)^2 \delta x^2 + \left(\frac{df}{dy}\right)^2 \delta y^2}$$

• Report $f \pm df$

• Examples: $f = x/y$
  $f = x + y$
  $f = \sin(x)$
Fitting

• Suppose we have dependent data vs. independent data (e.g. resistance vs. temperature)
• Can propose a function $f(x)$ to “fit” the data
• If data is $y_i$ at $x_i$ we want $f(x_i) \approx y_i$
• Use “chi-square” $\chi^2$ as a metric for “closeness”
Fitting (cont.)

• We choose a function that most closely represents what we think data is doing

• Create a single number $\chi^2$

\[ \chi^2 = \sum_{i=1}^{N} \left( \frac{f(x_i) - y_i}{\sigma_i} \right)^2 \]

• If each point is normally distributed, uncertainties are reasonably chosen, and fit function is good, we expect $\chi^2 \approx N$
Question

• How can you characterize this fit?
Answer

• It is suspiciously good, too good
• $\chi^2$ will be small
• Actually, it *may* be true and everything *may* have been done correctly
• Just not a likely circumstance
• The uncertainties seem to be too big
• Maybe too conservative an estimate for error
Question

• How can you characterize this fit?
Answer

• Fake data was generate with exact model
• About $\frac{1}{3}$ of data points don’t touch fit line
• This is expected for normally distributed data
• We also expect $\chi^2 \approx N$
Question

• How can you characterize this fit?
Answer

• Doesn’t look good, many points off fit line
• Expect $\chi^2$ to be large
• What can be gleaned from large $\chi^2$?
• Could be that uncertainties are too small, fit function is bad, or both
• This where $\chi^2$ can be bad at choosing
• What if uncertainties aren’t normal? Poisson?
More Characteristics of a Fit

• Fit function should have free parameters
• More free parameters generally makes fits better
• Maybe NOT physical, we must AVOID indiscriminately adding free parameters to fit unless absolutely necessary
• In fact, $\chi^2 \approx N - \nu$ where $\nu$ is the degrees of freedom
Question

• What are the degrees of freedom
  a.) 20 data points fit to a line
  b.) 15 data points fit to a sine wave with unknown frequency and amplitude, but fixed phase
  c.) 25 points fit to a completely unknown Gaussian + constant background
Answers

• a.) a line has $y = mx + b$ – this 2 free parameters – d.o.f. = $20 - 2 = 18$

b.) $f(x) = A\sin(kx + \phi)$ has 2 free parameters ($\phi$ is fixed) – d.o.f. = $15 - 2 = 13$

c.) Gaussian + constant background

$= y_0 + A \exp\{- (x-x_0)^2 / 2\sigma^2\}$ – d.o.f. = $25 - 4 = 21$
How a Fit Works

• Given a list of free parameters with possible bounds and constraints
• Walk through parameter “phase space”
• Calculate $\chi^2$ and its derivatives at each point
• Choose a direction to evolve your point selection
• Parameter search is a rich numerical problem