Physics 509: Error Propagation, and the Meaning of Error Bars

Scott Oser
Lecture #10

$Y = -3.50 + 10.9X$
$r = 0.539$
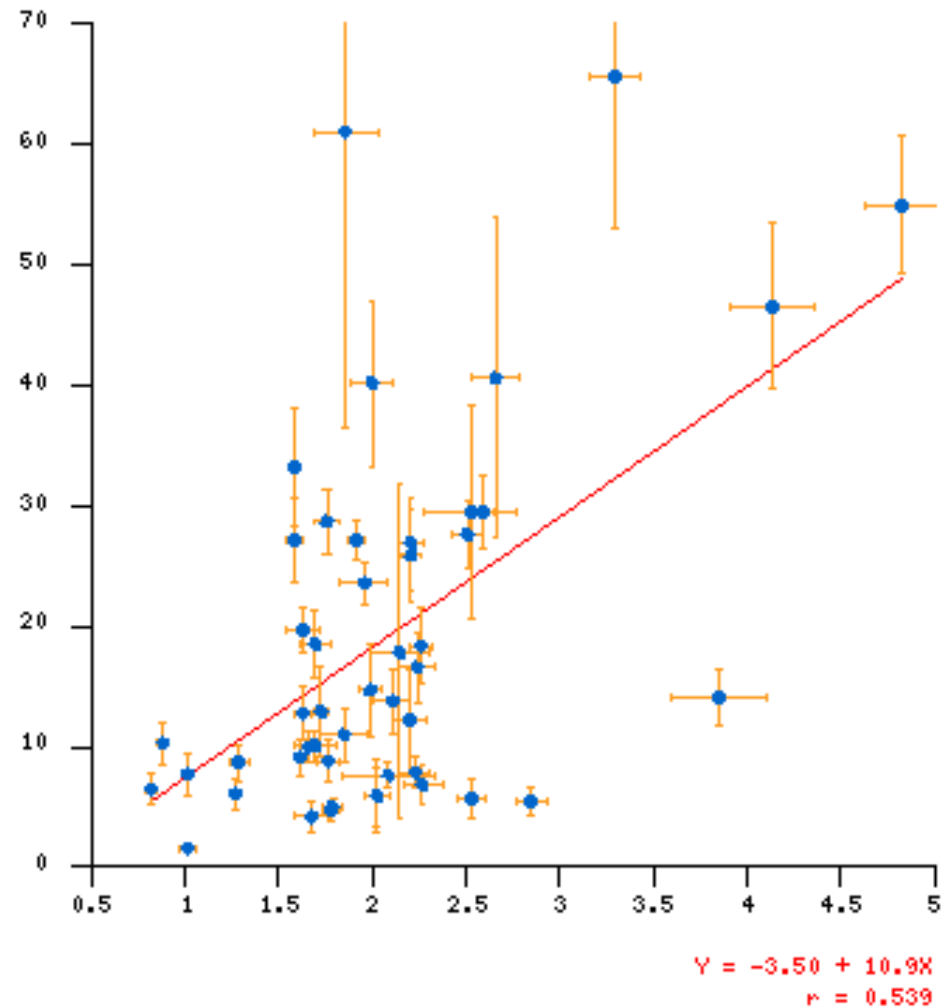
What is an error bar?

Y = -3.50 + 10.9X
r = 0.539
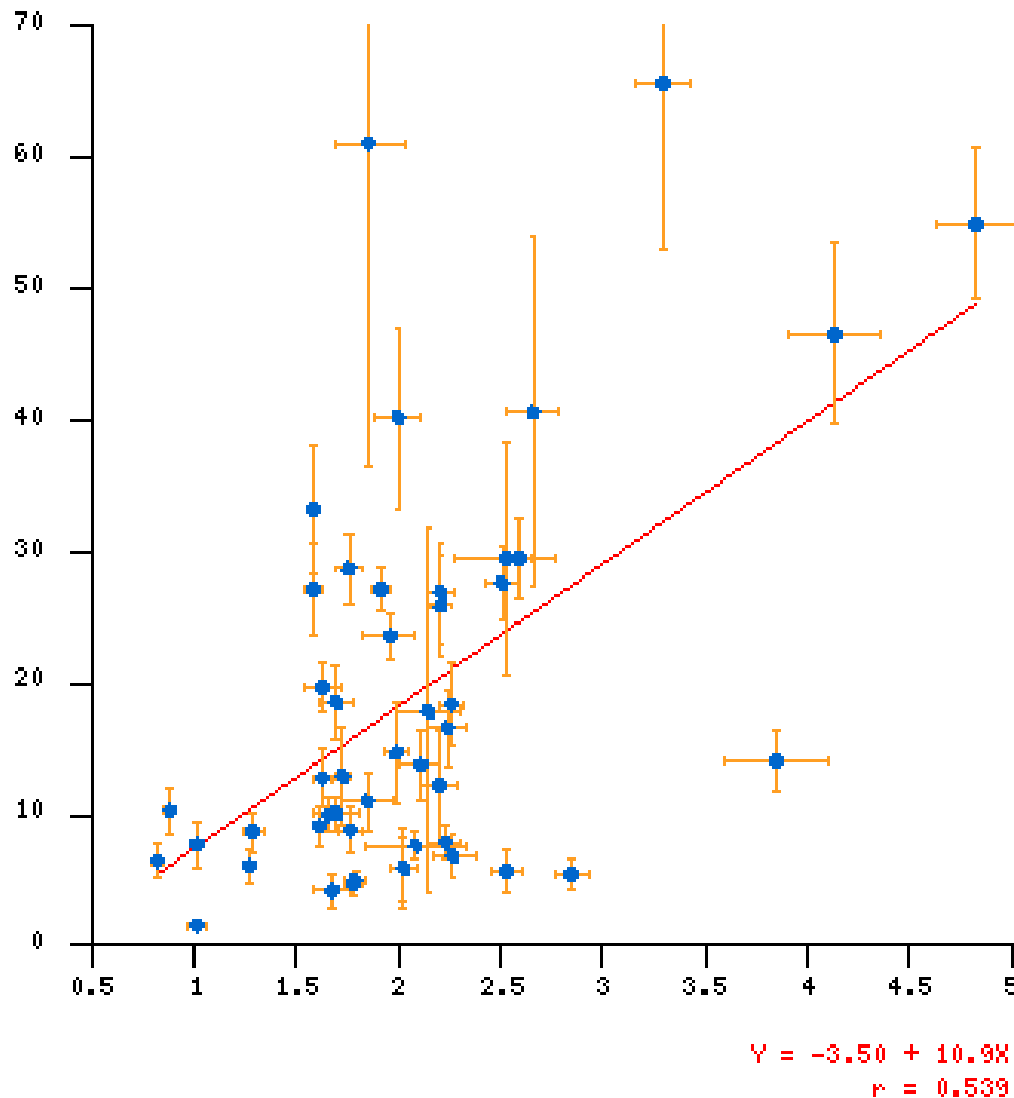
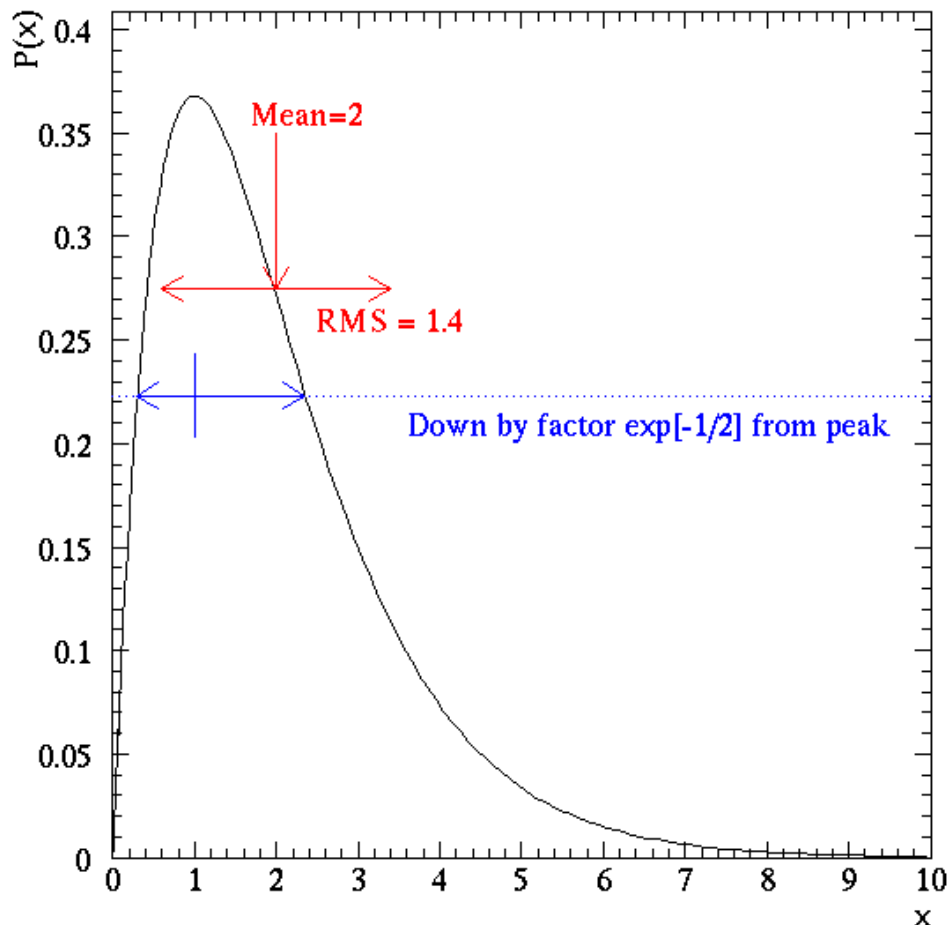Someone hands you a plot like this. What do the error bars indicate?

Answer: you can never be sure, unless it's specified!

Most common: vertical error bars indicate "±1σ" uncertainties.
Horizontal error bars can indicate uncertainty on X coordinate, or can indicate binning.

Correlations unknown!

2

Relation of an error bar to PDF shape

The error bar on a plot is most often meant to represent the ±1σ uncertainty on a data point. Bayesians and frequentists will disagree on what that means.

If data is distributed normally around "true value", it's clear what is intended:

$$\exp[-(x-\mu)^2/2\sigma^2].$$

But for asymmetric distributions, different things are sometimes meant ...

Mean=2

RMS = 1.4

Down by factor exp[-1/2] from peak

An error bar is a shorthand approximation to a PDF!

In an ideal Bayesian universe, error bars don't exist. Instead, everyone will use the full prior PDF and the data to calculate the posterior PDF, and then report the shape of that PDF (preferably as a graph or table).

An error bar is really a shorthand way to parameterize a PDF. Most often this means pretending the PDF is Gaussian and reporting its mean and RMS.

Many sins with error bars come from assuming Gaussian distributions when there aren't any.

An error bar as a confidence interval

Frequentist techniques don't directly answer the question of what the probability is for a parameter to have a particular value. All you can calculate is the probability of observing your data given a value of the parameter.The confidence interval construction is a dodge to get around this.
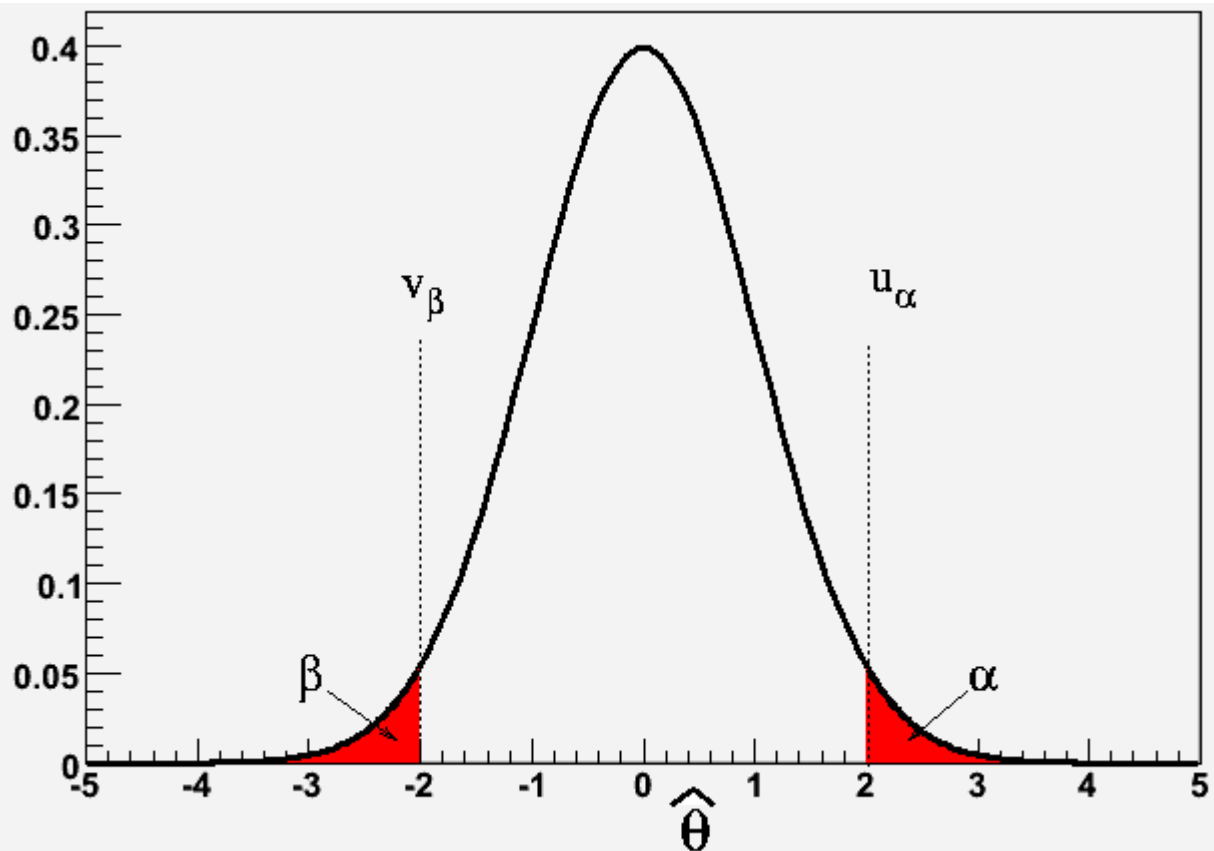
Starting point is the PDF for the estimator, for a fixed value of the parameter.

The estimator has probability $1-\alpha-\beta$ to fall in the white region.


5

Confidence interval construction
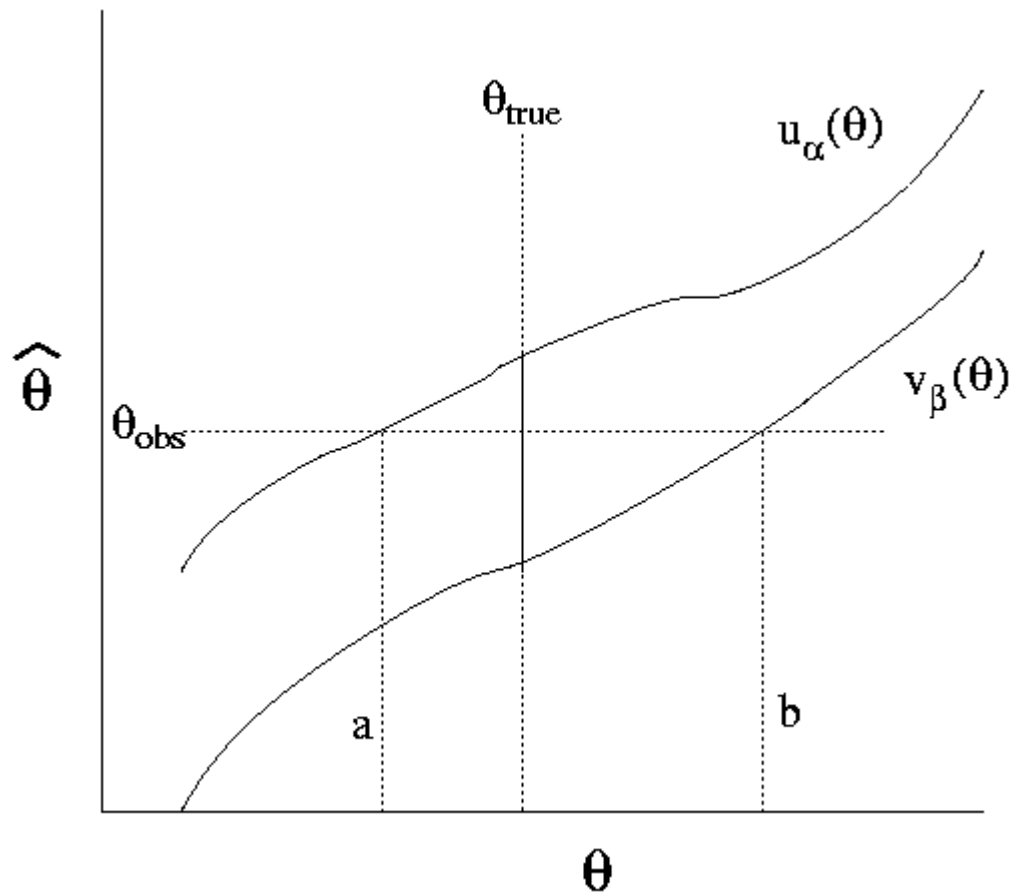
The confidence band is constructed so that the average probability of θ_{true} lying in the confidence interval is $1-\alpha-\beta$.

This should not be interpreted to mean that there is, for example, a 90% chance that the true value of θ is in the interval (a,b) on any given trial.

Rather, it means that if you ran the experiment 1000 times, and produced 1000 confidence intervals, then in 90% of these hypothetical experiments (a,b) would contain the true value.

Obviously a very roundabout way of speaking ...

The Δ ln(L) rule
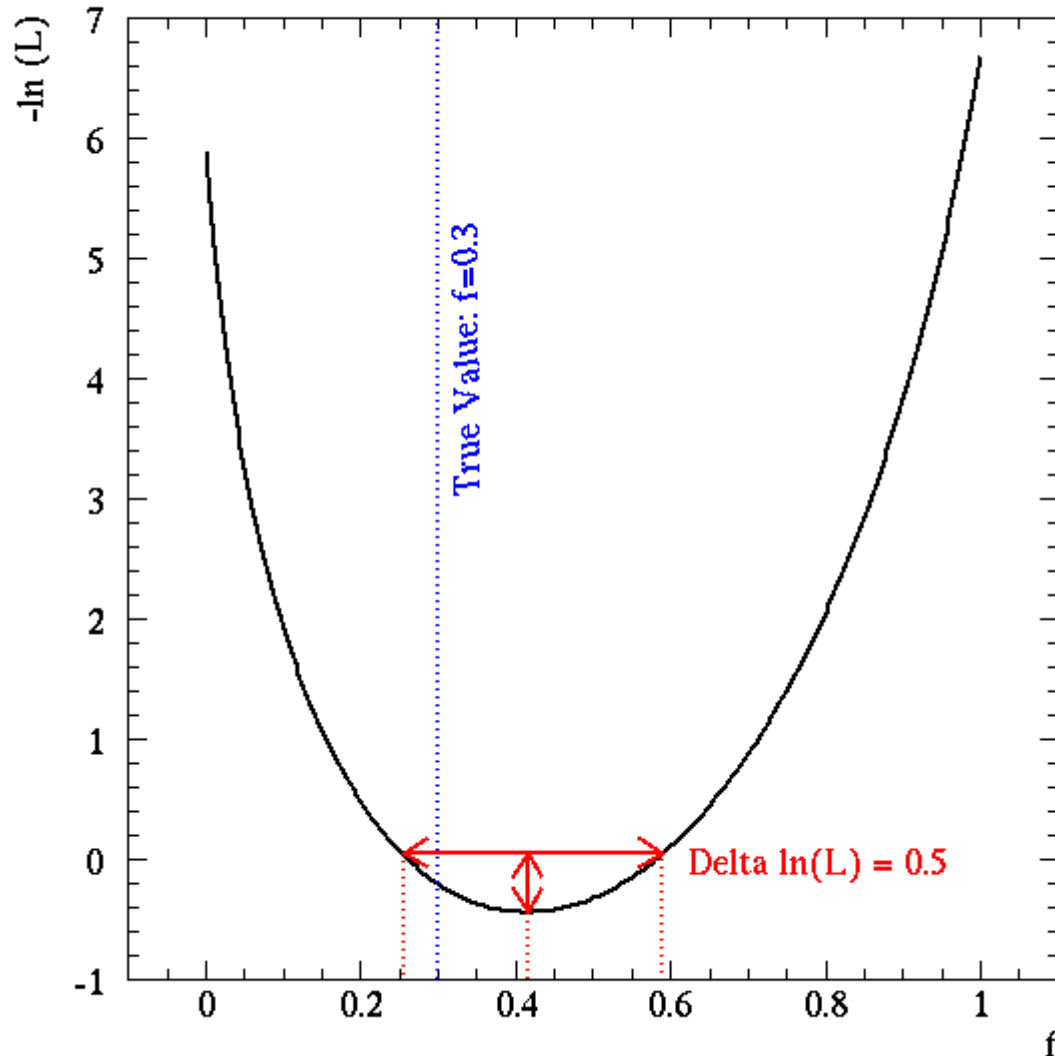
It is not trivial to construct proper frequentist confidence intervals. Most often an approximation is used: the confidence interval for a single parameter is defined as the range in which $\ln(L_{max})-\ln(L)<0.5$

This is only an approximation, and does not give exactly the right coverage when N is small.

More generally, if you have d free parameters, then the quantity $\omega = "\Delta\chi^2" = 2[\ln(L_{max})-\ln(L)]$ approximates a χ^2 with d degrees of freedom.

For experts: there do exist corrections to the Δ ln(L) rule that more accurately approximate coverage---see "Bartlett's correction". Often MC is better way to go.

Physics 509

7

Error-weighted averages

Suppose you have N independent measurements of a quantity. You average them. The proper error-weighted average is:

$$\langle x \rangle = \frac{\sum x_i / \sigma_i^2}{\sum 1/\sigma_i^2}$$

$$V(\langle x \rangle) = \frac{1}{\sum 1/\sigma_i^2}$$

If all of the uncertainties are equal, then this reduces to the simple arithmetic mean, with V(<x>) = V(x)/N.

Bayesian derivation of error-weighted averages

Suppose you have N independent measurements of a quantity, distributed around the true value μ with Gaussian distributions. For flat prior on μ we get:

$$P(\mu|\vec{x}) \propto \prod_i \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right] = \exp\left[-\frac{1}{2}\sum_i\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right]$$

It's easy to see that this has the form of a Gaussian. To find its peak, set derivative with respect to μ equal to zero.

$$\frac{dP}{d\mu} = \exp\left[-\frac{1}{2}\sum_i\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right]\left[\sum_i\left(\frac{x_i-\mu}{\sigma_i^2}\right)\right] = 0 \quad \rightarrow \quad \mu = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2}$$

Calculating the coefficient of μ^2 in the exponent yields:

$$V(\langle x \rangle) = \frac{1}{\sum 1/\sigma_i^2}$$

Physics 509

9

Averaging correlated measurements

We already saw how to average N independent measurement. What if there are correlations among measurements?

For the case of uncorrelated Gaussianly distributed measurements, finding the best fit value was equivalent to minimizing the chi-squared:

$$\chi^2 = \sum_i \left(\frac{x_i - \mu}{\sigma_i} \right)^2$$

In Bayesian language, this comes about because the PDF for μ is $\exp(-\chi^2/2)$. Because we know that this PDF must be Gaussian:

$$P(\mu) \propto \exp\left[-\frac{1}{2} \left(\frac{\mu - \mu_0}{\sigma_\mu} \right)^2 \right]$$

then an easy way to find the 1σ uncertainties on μ is to find the values of μ for which $\chi^2 = \chi^2_{min} + 1$.

Averaging correlated measurements II

The obvious generalization for correlated uncertainties is to form the χ^2 including the covariance matrix:

$$\chi^2 = \sum_i \sum_j (x_i - \mu)(x_j - \mu)(V^{-1})_{ij}$$

We find the best value of μ by minimizing this χ^2 and can then find the 1σ uncertainties on μ by finding the values of μ for which $\chi^2 = \chi^2_{min} + 1$.

This is really parameter estimation with one variable.

The best-fit value is easy enough to find:

$$\mu = \frac{\sum_{i,j} x_j (V^{-1})_{ij}}{\sum_{i,j} (V^{-1})_{ij}}$$

Averaging correlated measurements III

Recognizing that the χ^2 really just is the argument of an exponential defining a Gaussian PDF for μ ...
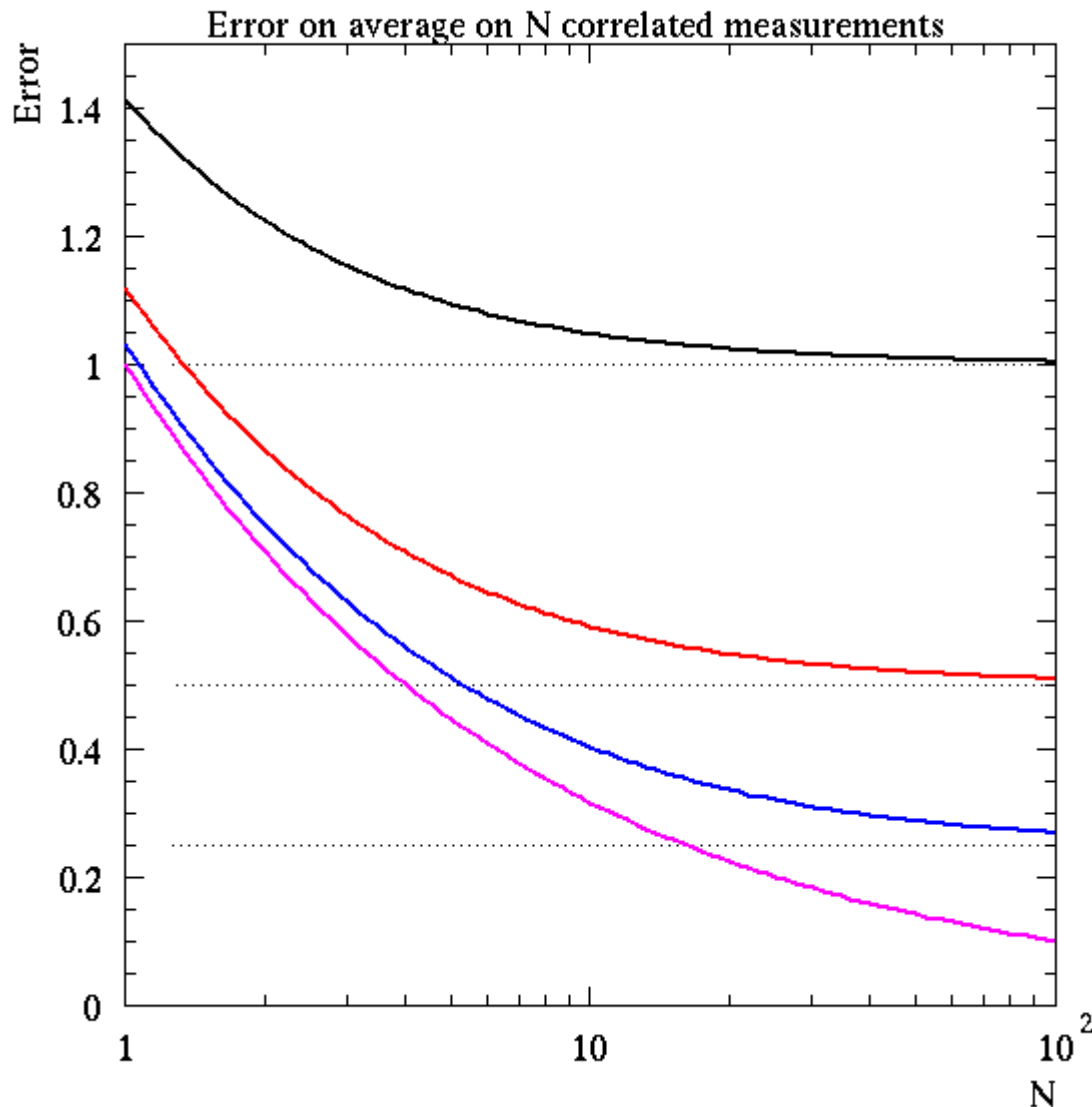
$$\chi^2 = \sum_i \sum_j (x_i - \mu)(x_j - \mu)(V^{-1})_{ij}$$

we can in fact read off the coefficient of μ^2, which will be $1/V(\mu)$:

$$\sigma^2_\mu = \frac{1}{\displaystyle\sum_{i,j} (V^{-1})_{ij}}$$

In general this can only be computed by inverting the matrix as far as I know.

Averaging correlated measurements IV

Error on average on N correlated measurements

Suppose that we have N correlated measurements. Each has some independent error σ=1 and a common error b that raises or lowers them all together. (You would simulate by first picking a random value for b, then for each measurement picking a new random value c with RMS σ and writing out b+c.

Each curve shows how the error on the average changes with N, for different values of b.

b=1
b=0.5
b=0.25
b=0

Averaging correlated measurements: example

Consider the following example, adapted from Cowan's book:

We measure an object's length with two rulers. Both are calibrated to be accurate at T=T_0, but otherwise have a temperature dependency: true length y is related to measured length by:

$$y_i = L_i + c_i (T - T_0)$$

We assume that we know the c_i and the uncertainties, which are Gaussian. We measure L_1, L_2, and T, and so calculate the object's true length y.

$$y_i = L_i + c_i (T - T_0)$$

We wish to combine the measurements from the two rulers to get our best estimate of the true length of the object.

Averaging correlated measurements: example

We start by forming the covariance matrix of the two measurements:

$$y_i = L_i + c_i(T - T_0)$$

$$\sigma_i^2 = \sigma_L^2 + c_i^2 \sigma_T^2$$

$$\text{cov}(y_1, y_2) = c_1 c_2 \sigma_T^2$$

We use the method previously described to calculate the weighted average for the following parameters:

$c_1 = 0.1$ $\qquad$ $L_1 = 2.0 \pm 0.1$ $\qquad$ $y_1 = 1.80 \pm 0.22$ $\qquad$ $T_0 = 25$
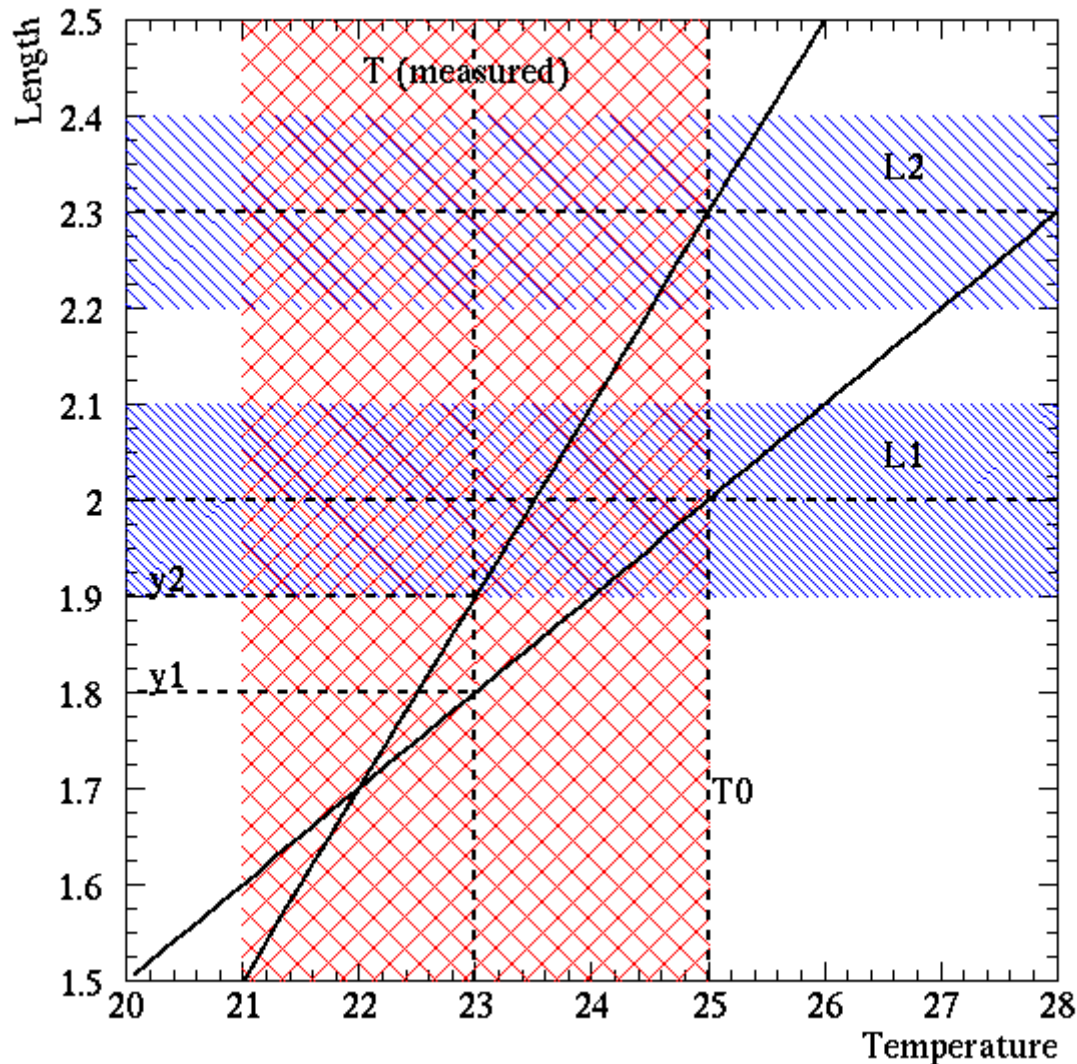$c_2 = 0.2$ $\qquad$ $L_2 = 2.3 \pm 0.1$ $\qquad$ $y_2 = 1.90 \pm 0.41$ $\qquad$ $T = 23 \pm 2$

Using the error propagation equations, we get for the weighted average:

$$y_{true} = 1.75 \pm 0.19$$

WEIRD: the weighted average is smaller than either measurement! What's going on??

Averaging correlated measurements: example

Bottom line:

Because y_1 and y_2 disagree, fit attempts to adjust temperature to make them agree. This pushes fitted length lower.

This is one of many cases in which the data itself gives you additional constraints on the value of a systematic (here, what is the true T).

The error propagation equation

Let *f(x,y)* be a function of two variables, and assume that the uncertainties on *x* and *y* are known and "small". Then:

$$\sigma_f^2 = \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 + 2\left(\frac{df}{dx}\right)\left(\frac{df}{dy}\right) \rho \, \sigma_x \sigma_y$$

The assumptions underlying the error propagation equation are:

• covariances are known
• *f* is an approximately linear function of *x* and *y* over the span of *x±dx* or *y±dy.*

The most common mistake in the world: ignoring the third term. Intro courses ignore its existence entirely!

Example: interpolating a straight line fit
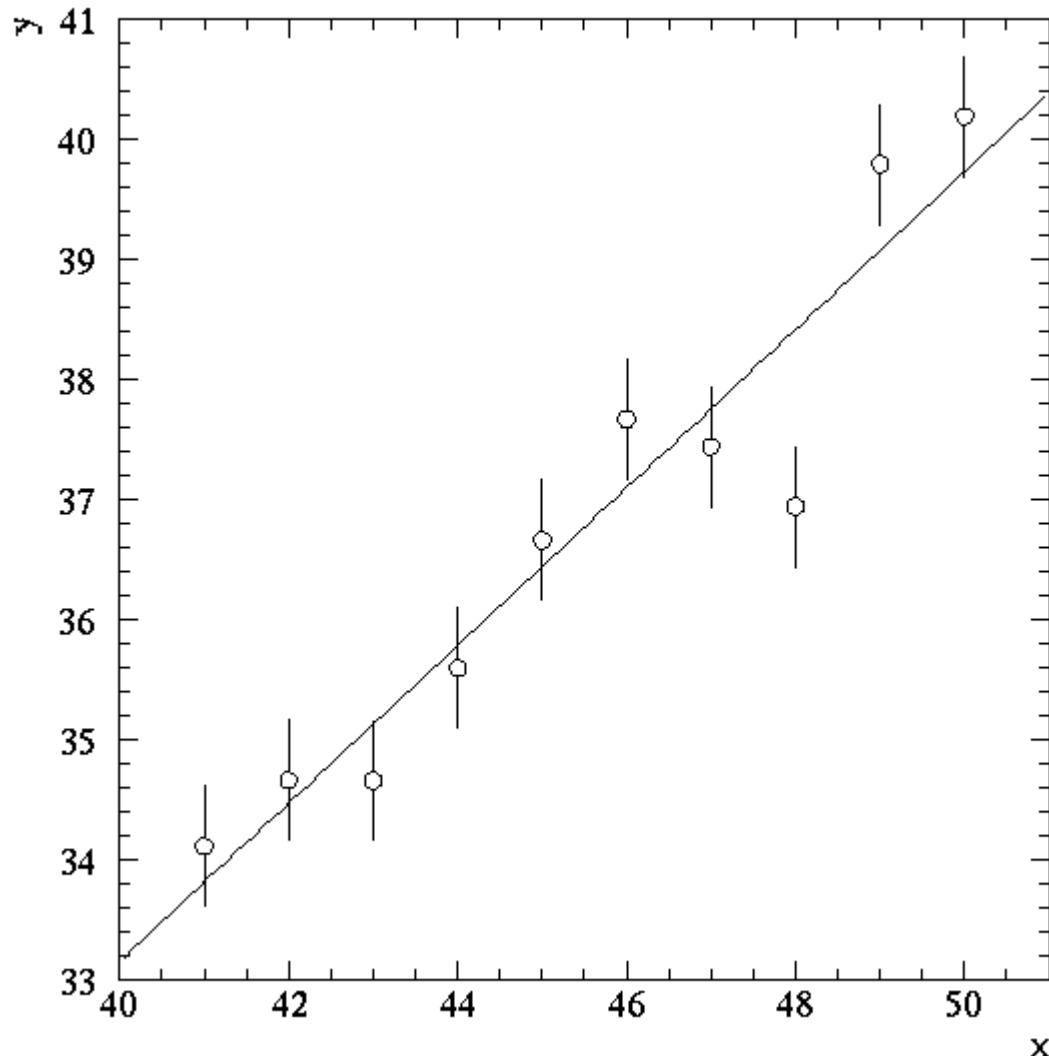
Straight line fit y=mx+b

Reported values from a standard fitting package:

m = 0.658 ± 0.056
b =   6.81 ± 2.57

Estimate the value and uncertainty of *y* when *x*=45.5:

y=0.658*45.5+6.81=36.75

$$dy=\sqrt{2.57^2+(45.5\cdot.056)^2}=3.62$$

UGH! NONSENSE!

Physics 509

18

Example: straight line fit, done correctly

Here's the correct way to estimate y at x=45.5. First, I find a better fitter, which reports the actual covariance matrix of the fit:

m = 0.0658 + .056
b = 6.81 + 2.57
ρ = -0.9981

$$dy = \sqrt{2.57^2 + (0.056 \cdot 45.5)^2 + 2(-0.9981)(0.056 \cdot 45.5)(2.57)} = 0.16$$

(Since the uncertainty on each individual data point was 0.5, and the fitting procedure effectively averages out their fluctuations, then we expect that we could predict the value of y in the meat of the distribution to better than 0.5.)

Food for thought: if the correlations matter so much, why don't most fitting programs report them routinely???

Reducing correlations in the straight line fit
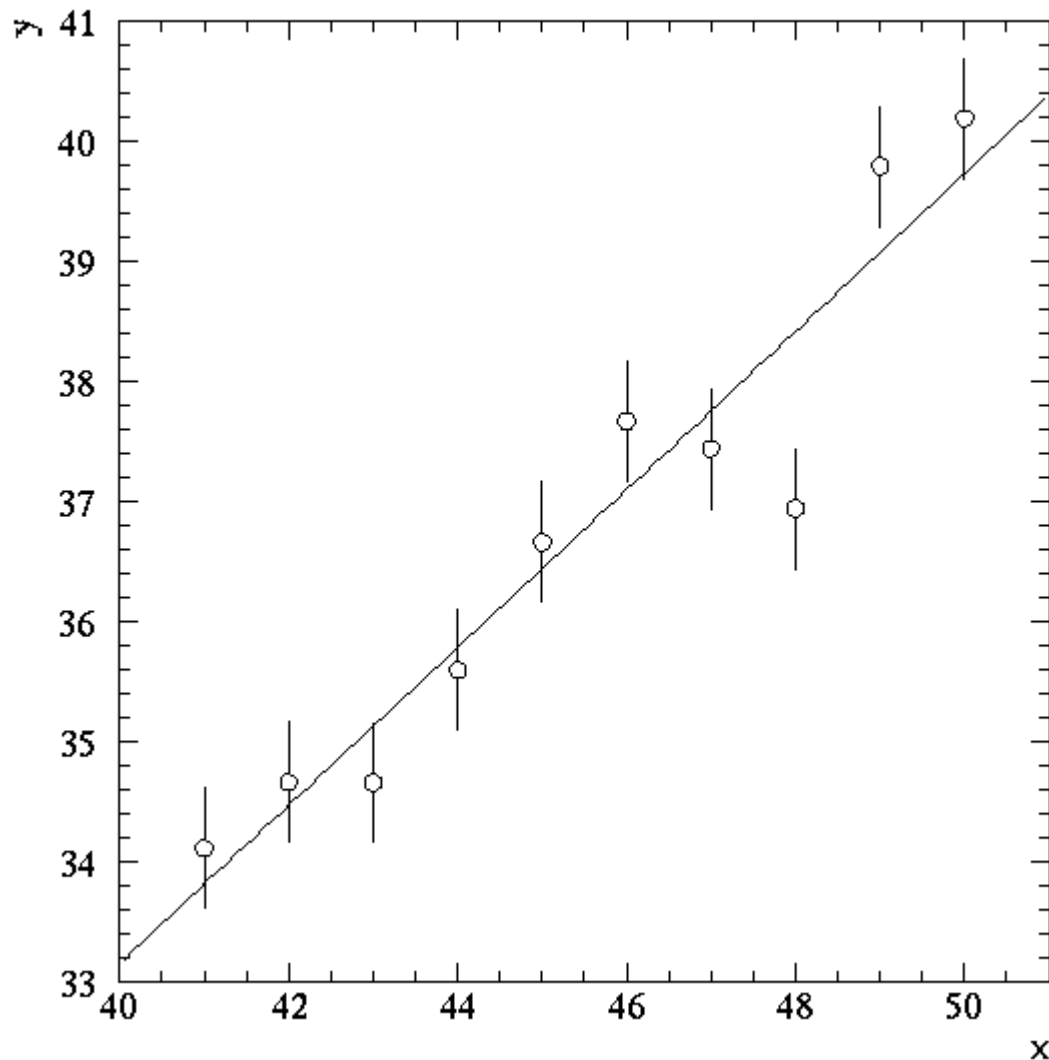
The strong correlation between m and b results from the long lever arm--- since you must extrapolate line to x=0 to determine b, a big error on m makes a big error on b.

You can avoid strong correlations by using more sensible parameterizations: for example, fit data to y=b'+m(x-45.5):

b' = 36.77 ± 0.16
m = 0.658 ± 0.085
ρ = 0.43

dy at x=45.5 = 0.16

Physics 509

20

Non-Gaussian errors

The error propagation can give the false impression that propagating errors is as simple as plugging in variances and covariances into the error propagation equation and then calculating an error on output.

However, a significant issue arises: although the error propagation equation is correct as far as it goes (small errors, linear approximations, etc), it is often not true that the resulting uncertainty has a Gaussian distribution! Reporting the central value and an RMS may be misleading.
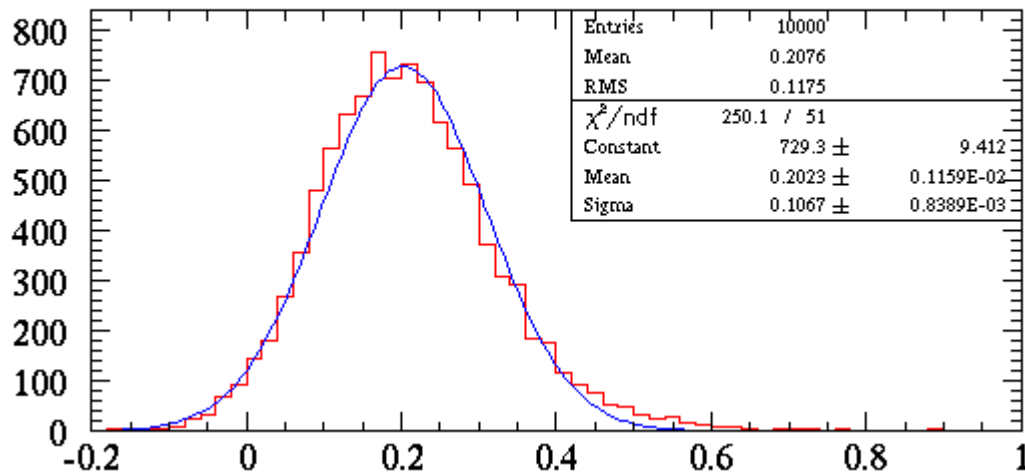
Ratio of two Gaussians I

Consider the ratio R=x/y of two independent variables drawn from normal distributions. By the error propagation equation

$$\left(\frac{dR}{R}\right)^2 = \left(\frac{dx}{x}\right)^2 + \left(\frac{dy}{y}\right)^2$$

Let's suppose x = 1 ± 0.5 and y = 5 ± 1. Then the calculated value of R = 0.200 ± .108.
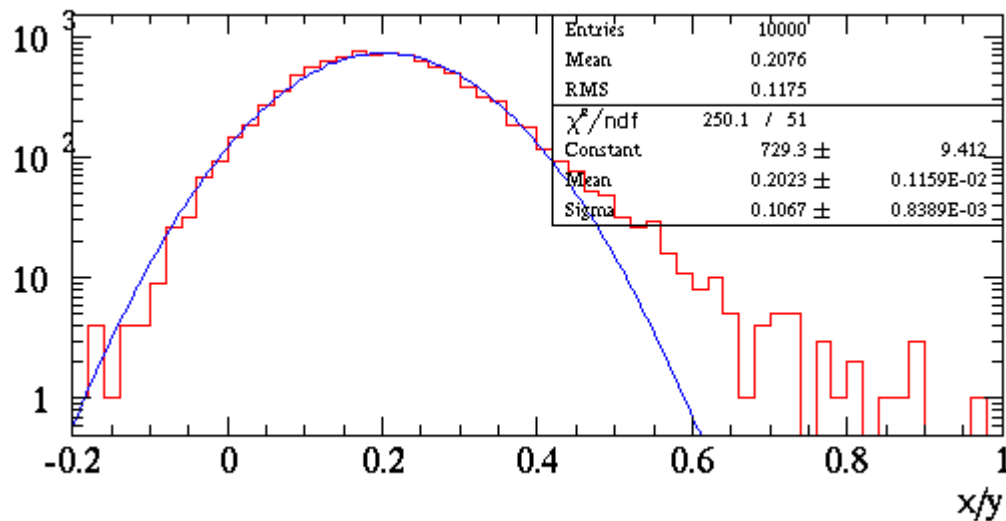
What does the actual distribution for R look like?

Ratio of two Gaussians II

$x = 1 \pm 0.5$ and $y = 5 \pm 1$.

Error propagation prediction:
$R = 0.200 \pm .108$.

Mean and RMS of R:
0.208 ± 0.118

Gaussian fit to peak:
0.202 ± 0.107

Non-Gaussian tails evident, especially towards larger R, but not too terrible.
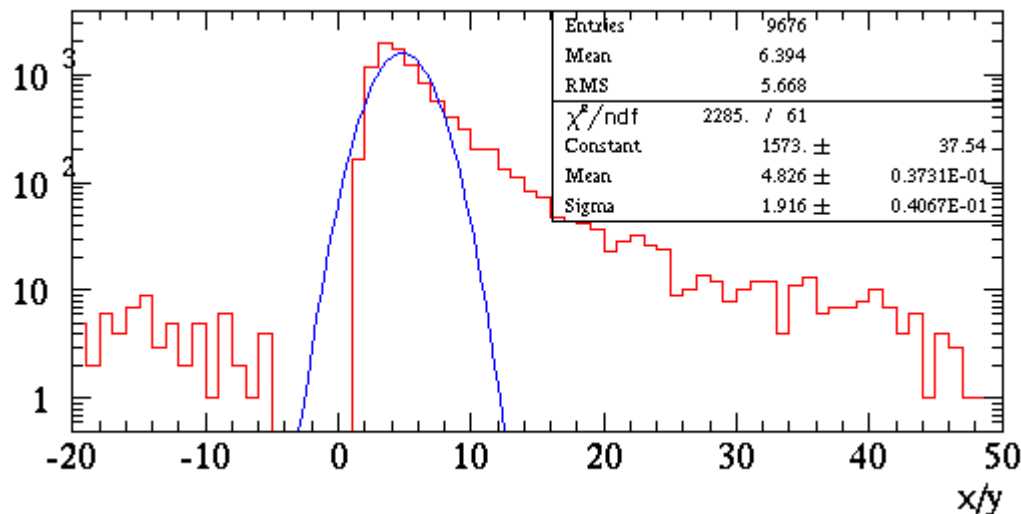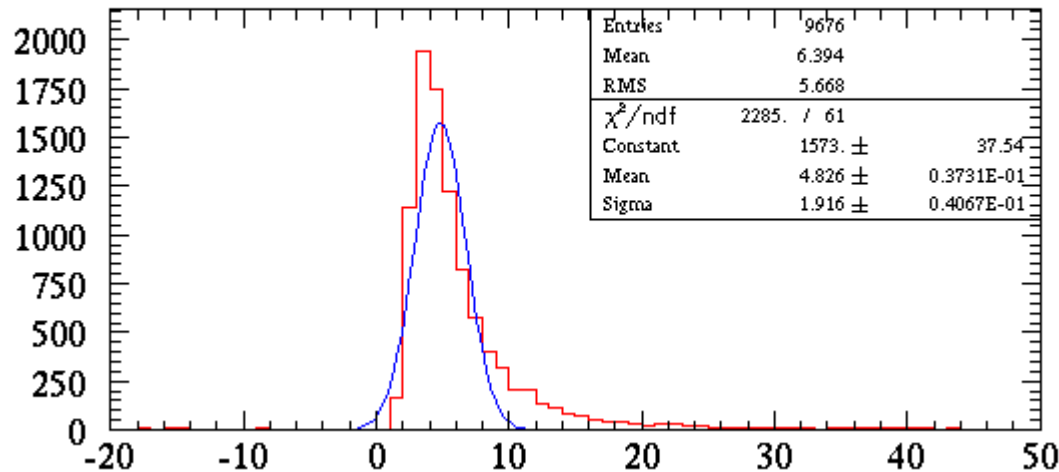
Ratio of two Gaussians III

$x = 5 \pm 1$ and $y = 1 \pm 0.5$

Error propagation:
$R = 5.00 \pm 2.69$.

Mean and RMS of R:
6.39 ± 5.67

Gaussian fit to peak:
4.83 ± 1.92

Completely non-Gaussian in all respects. Occasionally we even divide by zero, or close to it!

Physics 509

Ratio of two Gaussians IV

$$x = 5 \pm 1 \text{ and } y = 5 \pm 1$$
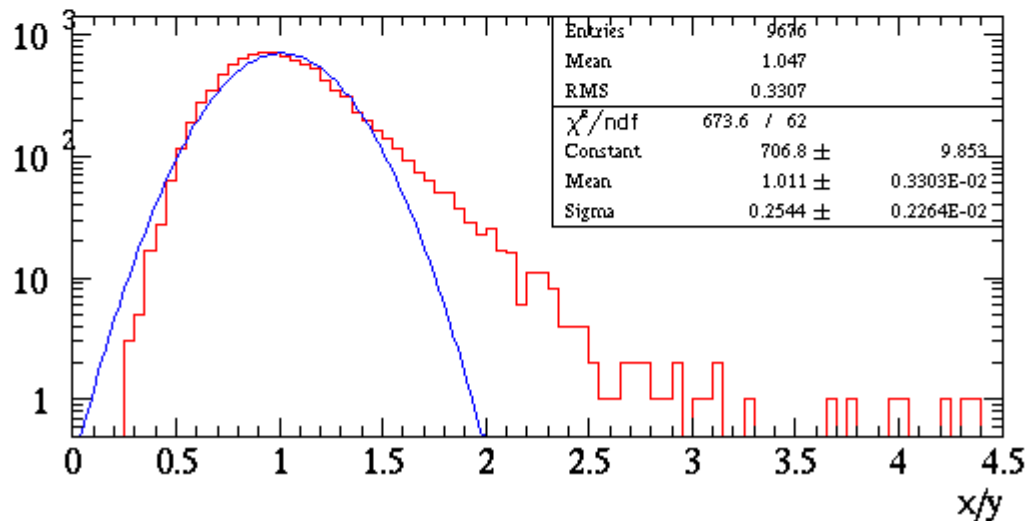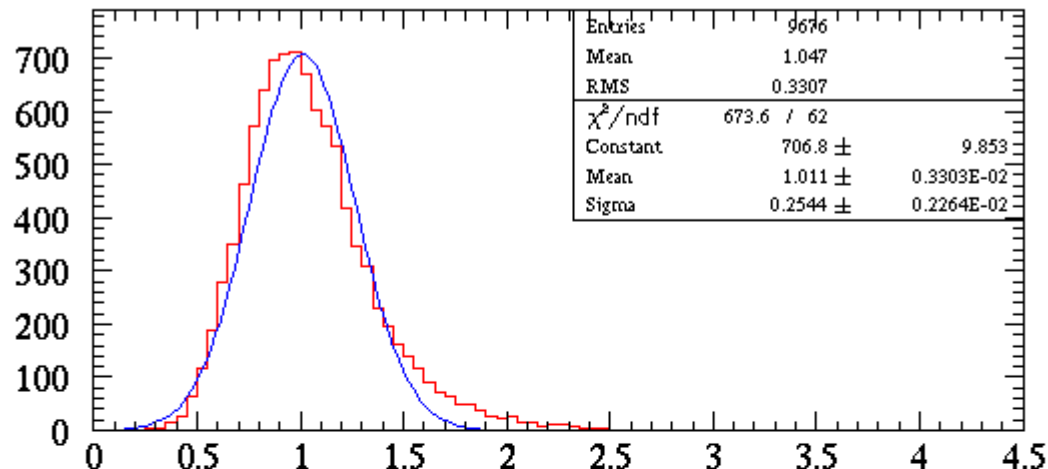
Error propagation:
R = 1 ± 0.28

Mean and RMS of R:
1.05 ± 0.33

Gaussian fit to peak:
1.01 ± 0.25

More non-Gaussian than first case, much better than second.

Rule of thumb: ratio of two Gaussians will be approximately Gaussian if fractional uncertainty is dominated by numerator, and denominator cannot be small compared to numerator.

Ratio of two Gaussians: testing an asymmetry

A word of caution: often scientists like to form asymmetries---for example, does the measurement from the left half of the apparatus agree with that from the right half? Asymmetry ratios are the usual way to do this, since common errors will cancel in ratio:

$$A = \frac{\phi_L - \phi_R}{\frac{1}{2}(\phi_L + \phi_R)}$$

Be extremely careful about using the error on the asymmetry as a measure of whether A is consistent with zero. In other words, A=0.5 ± 0.25 is usually not a "2 sigma" result, probability-wise.

Instead, it's better to simply test whether the numerator is consistent with zero or not.

Generalizing the error propagation equation

If we have N functions of M variables, we can calculate their covariance by:

$$\operatorname{cov}(f_k, f_l) = \sum_i \sum_j \left(\frac{\partial f_k}{\partial x_i}\right)\left(\frac{\partial f_l}{\partial x_j}\right) \operatorname{cov}(x_i, x_j)$$

We can write this compactly in matrix form as:

$$\boldsymbol{G} = G_{ki} = \left(\frac{\partial f_k}{\partial x_i}\right) \qquad \text{(an N X M matrix)}$$

$$V_f = \boldsymbol{G} \cdot V_x \cdot \boldsymbol{G}^T$$

Asymmetric errors

The error propagation equation works with covariances. But the confidence interval interpretation of error bars often reports asymmetric errors. We can't use the error propagation equation to combine asymmetric errors. What do we do?

Quite honestly, the typical physicist doesn't have a clue. The most common procedure is to separately add the negative error bars and the positive error bars in quadrature:

Source	- Error	+ Error
Error A	-0.025	+0.050
Error B	-0.015	+0.010
Error C	-0.040	+0.040
Combined	-0.049	+0.065

Warning: in spite of how common this procedure is and what you may have heard, it has no statistical justification and often gives the wrong answer!

How to handle asymmetric errors

Best case: you know the likelihood function (at least numerically). Report it, and include it in your fits.

More typical: all you know are a central value, with +/- errors. Your only hope is to come up with some parameterization of the likelihood that works well. This is a black art, and there is no generally satisfactory solution.
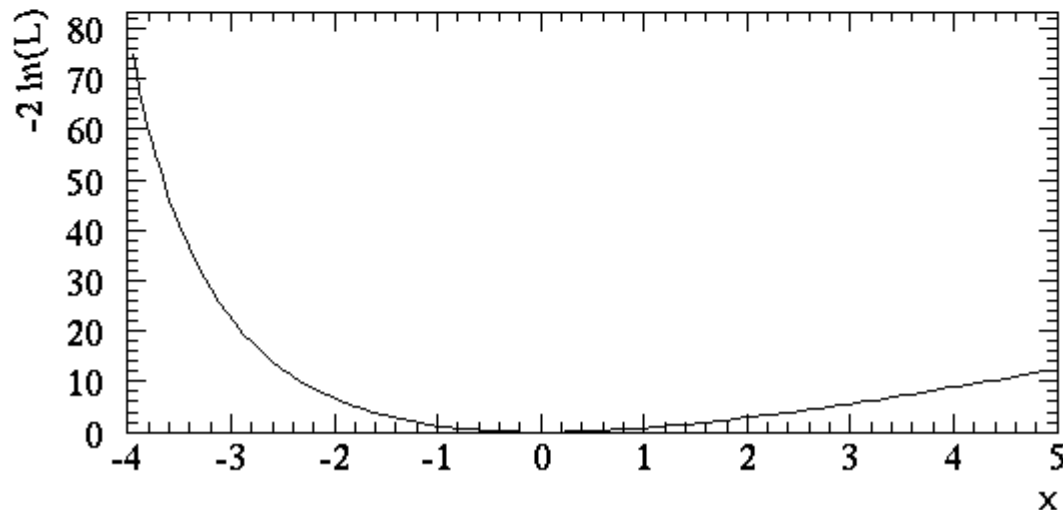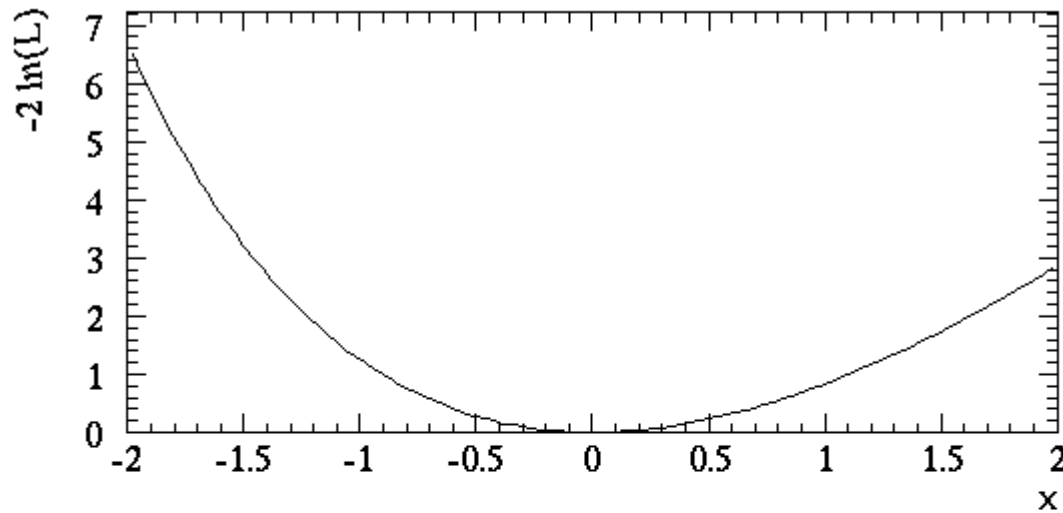
Barlow recommends (in a paper on his web site):

$$-\ln L(\alpha) = \frac{1}{2} \frac{\alpha^2}{\sigma_1 \sigma_2 + (\sigma_1 - \sigma_2) \alpha}$$

You can verify that this expression evaluates to ½ when $\alpha=+\sigma_1$ or $\alpha=-\sigma_2$

Parameterizing the asymmetric likelihood function

$$-\ln L(\alpha) = \frac{1}{2} \frac{\alpha^2}{\sigma_1 \sigma_2 + (\sigma_1 - \sigma_2)\alpha}$$

Note that parameterization breaks down when denominator becomes negative. It's like saying there is a hard limit on α at some point. Use this only over its appropriate range of validity.

Barlow recommends this form because it worked well for a variety of sample problems he tried---it's completely empirical.

See Barlow, "Asymmetric Statistical Errors", arXiv:physics/0406120

Application: adding two measurements with asymmetric errors.

Suppose $A = 5^{+1}_{-2}$ and $B = 3^{+1}_{-3}$. What is the confidence interval for A+B?

$$-\ln L(A,B) = \frac{1}{2} \frac{(A-5)^2}{(1)(2)+(1-2)(A-5)} + \frac{1}{2} \frac{(B-3)^2}{(1)(3)+(1-3)(B-3)}$$

Let C=A+B. Rewrite likelihood to be a function of C and one of the other variables---eg.

$$-\ln L(C,A) = \frac{1}{2} \frac{(A-5)^2}{(1)(2)+(1-2)(A-5)} + \frac{1}{2} \frac{(C-A-3)^2}{(1)(3)+(1-3)(C-A-3)}$$

To get the likelihood function for C alone, which is what we care about, minimize with respect to nuisance parameter A:

$$-\ln L(C) = \min_{A} \; -\ln L(C,A)$$

Asymmetric error results