

Physics 509: Intro to Systematic Errors

Scott Oser
Lecture #11
October 16, 2008



What does Lincoln Wolfenstein have to do with systematics?

Lincoln Wolfenstein---professor emeritus at Carnegie Mellon. Famous as the W in the MSW effect (matter effects in neutrino oscillation), and for a variety of contributions to particle physics.

When told that the SNO collaboration was not ready to release its first results because we were finalizing our systematic uncertainties, he said:

“Systematics ... that's when you guys just vote, right?”



Was he implying that systematic error assignments are less than rigorous? ☺

More right than you might care to think ...

A possibly apocryphal story ... since I can't prove it's true, I will omit the names of the people in question. (But buy me a beer, and I'll tell you.)

A large particle physics experiment was arguing vigorously about how large a particular systematic uncertainty was. After several hours of discussion failed to reach agreement, the spokesperson said:

“OK, everyone who thinks the systematic is smaller than 0.5%, raise your hand.”

“Now, those people keep your hands up. Anyone who thinks the systematic uncertainty is smaller than 1% should also raise your hand.”

“2%? 2.5%? 3%?”

The spokesperson took the value at which 50% of the collaboration had their hand in the air, and told everyone to use that value as the systematic.

Discussion: Was this crazy?

What is a systematic uncertainty?

There are many meanings of the term “systematic uncertainty”. (I prefer this term to “systematic error”, which means more or less the same thing.)

The most common definition is “any error that's not a statistical error”.

To avoid this definition becoming circular, we'd better be more precise.

Perhaps this works: “A systematic uncertainty is a possible unknown variation in a measurement, or in a quantity derived from a set of measurements, that does not randomly vary from data point to data point.”

Usually you see it listed broken out as: 5.0 ± 1.2 (stat) ± 0.8 (sys)

Examples of systematic uncertainties

“Like sands through an hourglass, so are the systematics of our lives ...”

- You measure the length of an object, but worry that the ruler might have contracted slightly due to it being a cold day.
- You try to infer the brightness of a distant supernova, but worry that intervening dust might make it seem dimmer than you expect.
- Your thermometer is miscalibrated.
- You measure $g-2$, the anomalous magnetic moment of the muon, and ask whether it agrees with the Standard Model expectation. A theorist tells you that there are higher order corrections to the theory prediction that are too complicated for her to calculate, but she helpfully quotes an uncertainty based on how large she believes these are likely to be.
- You are trying to fit an energy spectrum to an expected shape plus a background component to determine the size of a signal. There are two experimental measurements of the expected shape. They disagree by an amount much larger than their error bars.

Why are systematics problematic for frequentists?

The whole frequentist program is based upon treating the outcomes of experiments as “random variables”, and predicting the probabilities of observing various outcomes. For quantities that fluctuate, this makes sense.

But often we conceive of systematic uncertainties that aren't fluctuations. Maybe your thermometer really IS off by 0.2K, and every time you repeat the measurement you'll have the same systematic bias.

There's both a conceptual problem and a practical problem here. Conceptually, we resort to the dodge of imagining “identical” hypothetical experiments, except that certain features of the setup are allowed to vary. Practically, we usually can't measure the size of a systematic by repeating the measurement 100 times and looking at the distribution. We're almost forced to be pseudo-Bayesian about the whole thing.

Bayesian approach to systematics

Bayesians lose no sleep over systematics. Suppose you want to measure some quantity θ . You have a prior $P(\theta|I)$, you observed some data D , and you need to calculate a likelihood $P(D|\theta,I)$. Let's suppose that the likelihood depends on some systematic parameter α (which could for example be the offset on our thermometer). We handle the systematic uncertainty by simply treating both θ and α as unknown parameters, assign a prior to each, and write down Bayes theorem:

$$P(\theta, \alpha|D, I) = \frac{P(D|\theta, \alpha, I) P(\theta, \alpha|I)}{\int d\theta d\alpha P(D|\theta, \alpha, I) P(\theta, \alpha|I)}$$

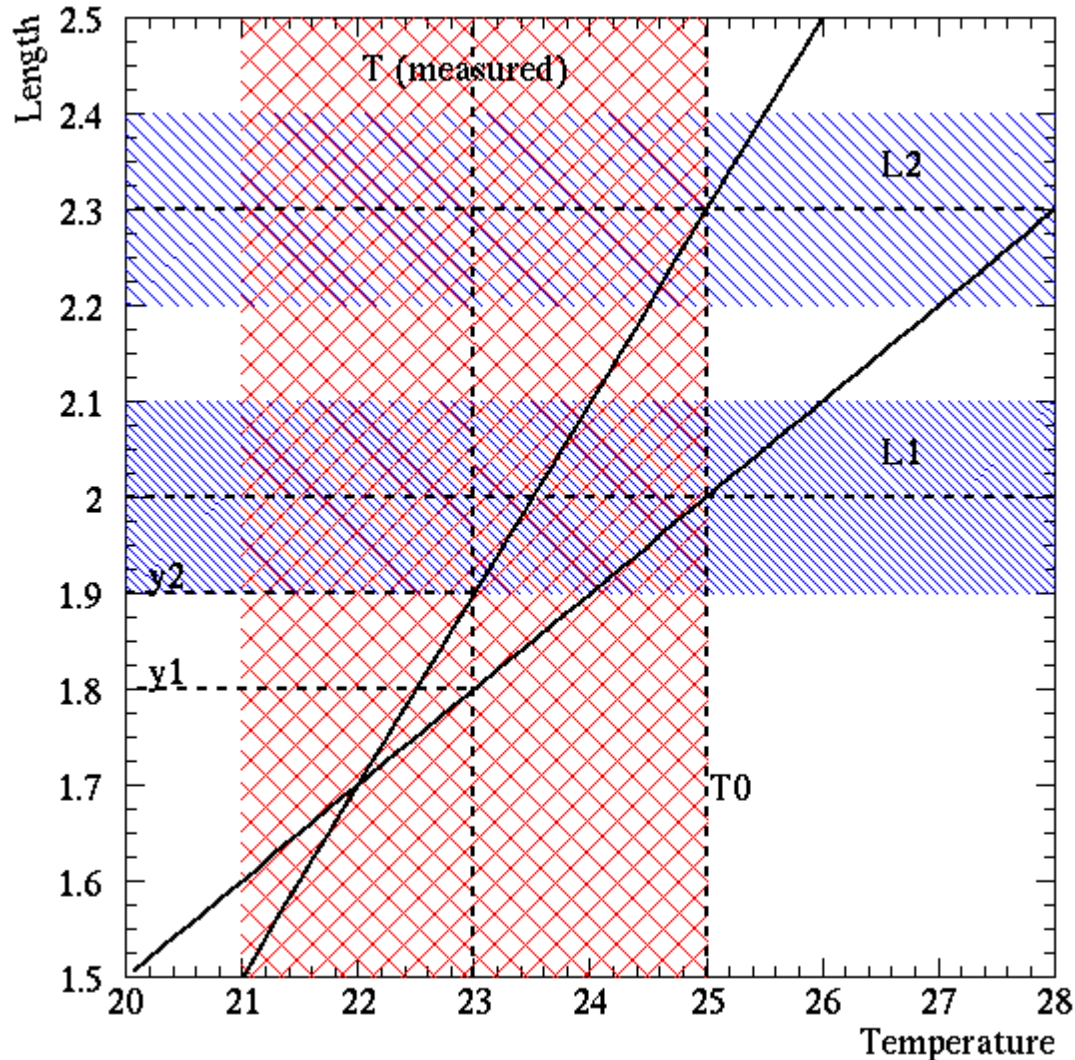
In the end we get a distribution for θ , whose value we care about, and for α , which may be uninteresting. We marginalize by integrating over α to get $P(\theta|I)$.

The prior $P(\alpha)$ presents our prior knowledge of α and is often the result of a calibration measurement.

Note that since the likelihood $P(D|\theta,\alpha,I)$ depends on α as well, it can provide additional information on α .

A case where the data constrained a systematic

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



$$\begin{array}{ll} c_1 = 0.1 & c_2 = 0.2 \\ L_1 = 2.0 \pm 0.1 & L_2 = 2.3 \pm 0.1 \\ y_1 = 1.80 \pm 0.22 & y_2 = 1.90 \pm 0.41 \\ T_0 = 25 & T = 23 \pm 2 \end{array}$$

The intersection of the two lines in this example from last class provides a better estimate of the true temperature than that provided from the external calibration of 23 ± 2 .

Distinction between statistical and systematic uncertainties

A common set of definitions:

A “statistical uncertainty” represents the scatter in a parameter estimation caused by fluctuations in the values of random variables. Typically this decreases in proportion to $1/\sqrt{N}$.

A “systematic uncertainty” represents a constant (not random) but unknown error whose size is independent of N .

DO NOT TAKE THESE DEFINITIONS TOO SERIOUSLY. Not all statistical uncertainties decrease like $1/\sqrt{N}$. And more commonly, taking more data can decrease a systematic uncertainty as well, especially when the systematic affects different parts of the data in different ways, as in the example on the previous page.

Need to have a systematics model

The most important step in dealing with any systematic is to have a quantitative model of how it affects the measurement. This includes:

- A. How does the systematic affect the measured data points themselves?
- B. How does the systematic appear quantitatively in the calculations applied to the data?

It is essential to have some model, however simplified, in order to quantify the systematic uncertainty.

Systematic error model #1: an offset

Suppose we take N measurements from a distribution, and wish to estimate the true mean of the underlying distribution.

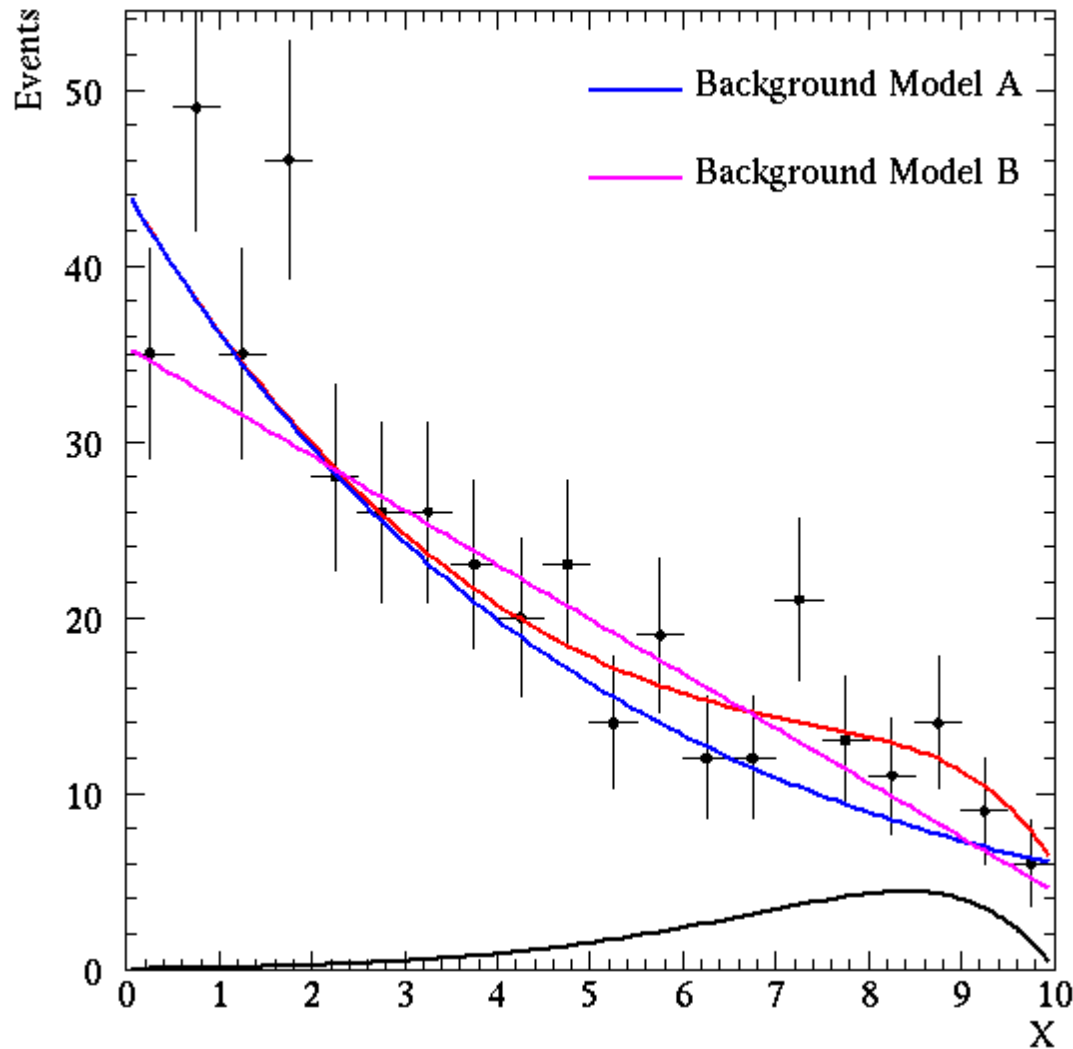
Our measuring apparatus might have an offset s from 0. We attempt to calibrate this. Our systematic error model consists of:

- 1) There is some additive offset s whose value is unknown.
- 2) It affects each measurement identically by $x_i \rightarrow x_i + s$.
- 3) The true mean is estimated by:

$$\hat{\mu} = \left(\frac{1}{N} \sum_{i=1}^N x_i \right) - s$$

- 4) Our calibration is $s = 2 \pm 0.4$

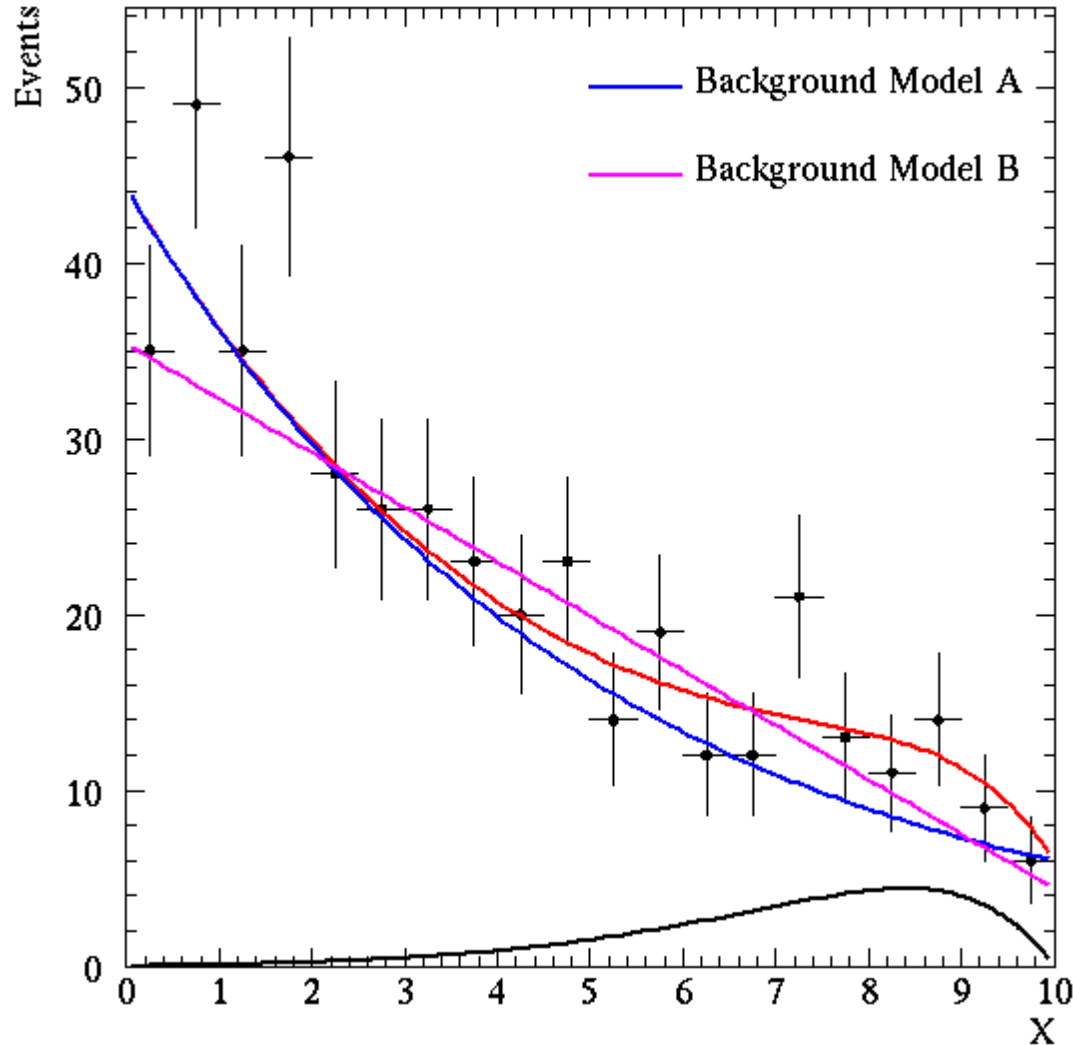
Systematic error model #2: two incompatible models



In order to determine the rate of some process, we fit the data to a two-component model consisting of a signal shape and a background shape.

But there are two different and mutually exclusive background models, which we'll denote as A and B.

Systematic error model #2: two incompatible models



If the error ranges on the background models are negligible, one possibility is to just do the analysis twice, reporting the result with each model, and hope that future information will determine which model is right.

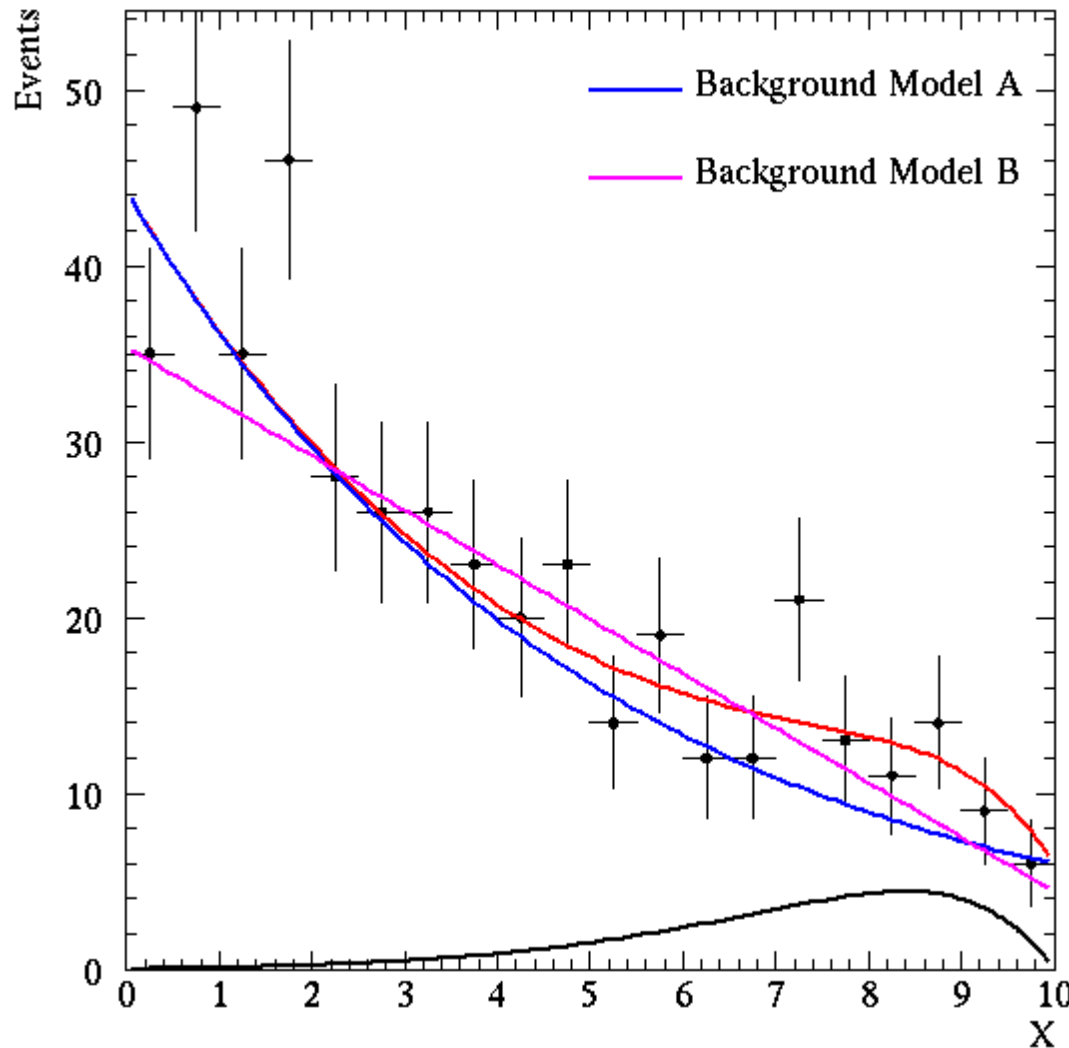
But in this case the shape of the data actually will tell us something about the two models---data will constrain the systematic (the shape of the background).

Systematic error model #2: two incompatible models

One approach is to make a parameterized background model that interpolates between the two:

$$m(x) = fm_A(x) + (1-f)m_B(x)$$

Here $0 \leq f \leq 1$. You can define whatever Bayesian prior you like for f (even Dirac delta functions at $f=0$ and $f=1$). Your fit to the data will favour some values of f and not others, but the most important thing is you've quantified the problem through nuisance parameter f .



How do you measure a systematic?

So you've quantified the effects of the systematic through some nuisance parameter. How do you determine the value of that nuisance parameter itself? Various approaches:

- 1) Calibration measurements, taken separately from your main data
- 2) A priori estimate based on known parameters of the apparatus
- 3) If data provides useful data about nuisance parameter, fit it from the main data itself.
- 4) “Theory”: some systematic uncertainties will be what we call “theoretical uncertainties”. There are various causes/interpretations:
 - A. Measurement uncertainties in theory parameters
 - B. Theorists' estimates of errors due to approximations made
 - C. Spread between different theory estimates (careful here!)
- 5) Data vs. Monte Carlo comparisons: use calibration data to estimate how well Monte Carlo reproduces data, then use spread as an estimate of how well Monte Carlo predicts other quantities

How do you measure a systematic?

This is a black art. I'd argue that 90% of experimental physics is thinking of clever ways to reduce or at least measure systematics.



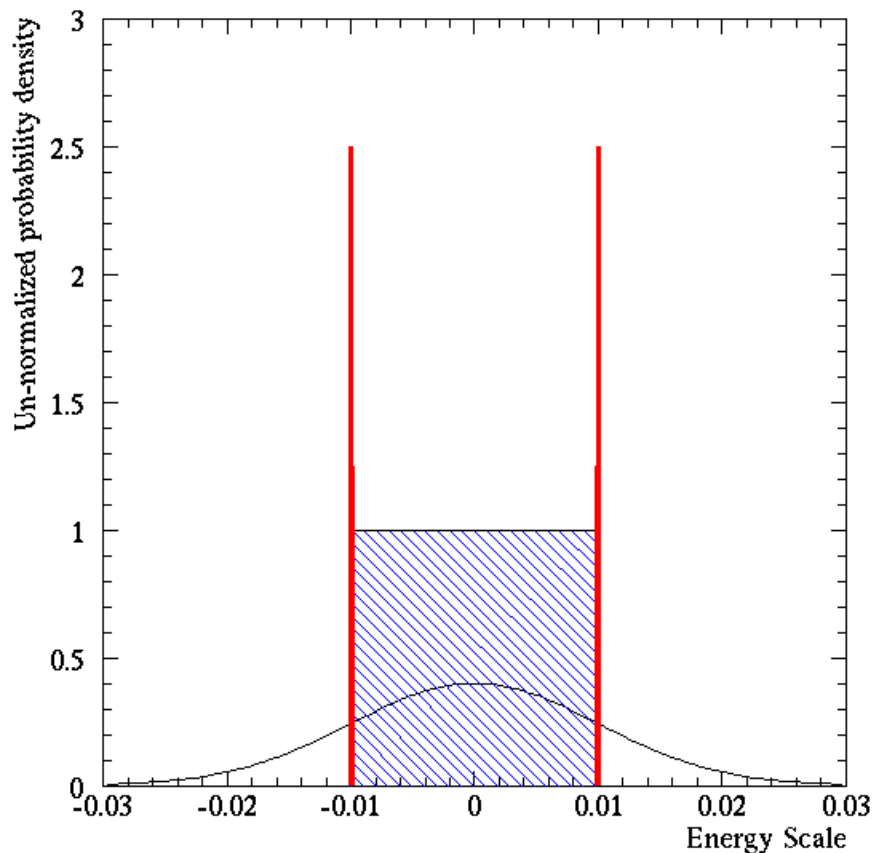
Severus Snape, dabbler in the black arts

Unfortunately, there is no real magic, merely hard work.

A strong dose of paranoia helps as well.

Avoid inflating systematics

There is a regrettable tendency to overestimate systematics in the name of CYA or to save effort. For example, perhaps you've concluded that the energy scale of your detector is at worst off by 1%. So you write $\sigma_E=0.01$, and proceed to treat this as the RMS of your nuisance parameter.



Black: typical Gaussian PDF implied by $\sigma=0.01$. Long tails far beyond “worst case” range.

Red: δ functions---only PDF with $\sigma=0.01$ that is fully contained within “worst case” range

Blue: maximum entropy distribution consistent with “worst case range. $\sigma=0.02/\sqrt{12}=0.0058$

Which should you use?

Why you should avoid inflating systematics...

What's wrong with inflating systematics to cover all bases? Isn't this the “conservative” thing to do?

- 1) This tends to paper over model inaccuracies, and imply greater support for your model than is warranted. (Think Bayesian-wise: since Bayesian analyses always choose *between* competing hypotheses, being “conservative” with one hypothesis is equivalent to selectively favouring another.
- 2) Your inflated error might hide a serious problem with your data, or worst of all may miss an important discovery.
- 3) Tendency to bias: everyone recognizes that it's wrong to “fudge your data” to make your central value agree with expectations. Fewer people recognize that it's equally wrong to inflate your errors to make sure the error bars overlap the expected value!

Propagating systematics with Monte Carlo

So you've listed all of the systematics, mapped them all to nuisance parameters (or decided that they're negligible), and have assigned PDFs to each nuisance parameter. What next?

“Propagating the systematics” means to determine how much uncertainty results in your final value from your systematics model. Toy Monte Carlo is an excellent way to do evaluate this:

- 1) Randomly choose values for each nuisance parameter according to their respective PDFs.
- 2) Analyze the data as if those values of the nuisance parameters are the true values for the systematic parameters.
- 3) Repeat many times.
- 4) If you're trying to estimate the error on a fit parameter, plot the distribution of the fitted values of that parameter. Take the RMS width as the systematic error.
- 5) If you're doing hypothesis testing, Monte Carlo both the systematics and the data, and plot the distribution of your test statistic to see how (un)likely it is to observe the given value.

Propagating systematics with Monte Carlo 2

Advantages of the Monte Carlo method:

- few approximations made---no need to assume Gaussian errors
- considers the effects of all systematics jointly, including nonlinearities
- can easily accommodate correlations between systematics

Disadvantages of the Monte Carlo method:

- method does not allow the data itself to constrain the systematics (although we will examine how to correct for this)
- because all systematics are varied at once, the resulting distribution is the convolution of the effects of all nuisance parameters. On the one hand this is a feature---in real life all systematics vary at once, and so Monte Carlo gives an “exact” way of modelling how various systematics interact. On the other hand, if you want to understand the relative importance of each component, you have to either marginalize or project over each parameter, or rerun your Monte Carlos, this time varying just one systematic at a time. (Actually, this is recommended practice in any case.)

Covariance matrix approach

Monte Carlo is not always necessary, and not always the fastest way to propagate systematics. In the “covariance matrix” approach, you treat the nuisance parameter s and the data values x_j as a set of correlated random variables. You then calculate their full covariance matrix, and use error propagation to estimate the uncertainties.

Ex. taking the average of a set of measurements with a systematic additive offset:

$$x_j = \mu + X_j + s$$

(Implicitly assuming X_j is independent of s).

$$\text{cov}(x_i, x_j) = \text{cov}(\mu + X_i + s, \mu + X_j + s) = \text{cov}(X_i, X_j) + \text{cov}(s, s)$$

You can think of this as the sum of two covariance matrices:

$$V_{\text{tot}} = V_{\text{stat}} + V_{\text{sys}}$$

Covariance matrix approach 2

Now just include the new covariance matrix in your analysis wherever you previously had just the statistical error covariances--- e.g.

$$\chi^2(\theta) = \sum_{i=1}^N \sum_{j=1}^N (y_i - f(x_i|\theta)) V_{ij}^{-1} (y_j - f(x_j|\theta))$$

Note: in this approach you often will consider the value of s to be fixed at its central value. In other words, although the covariance matrix V will contain information on how much the uncertainties on the measured values y_i are increased by the systematic, the above formulation doesn't directly yield a refined estimate of s . We'll correct this shortly.

Constraint terms in the likelihood

Working in Bayesian language, the posterior PDF is given by

$$P(\theta, \alpha | D, I) \propto P(\theta | I) P(\alpha | I) P(D | \theta, \alpha, I)$$

We saw previously that the ML estimator is same thing as the mode of the Bayesian posterior PDF assuming a flat prior on θ . In that case we maximized $\ln L(\theta) = \ln P(D | \theta, I)$, and use the shape of $\ln L$ to determine the confidence interval on θ .

This easily generalizes to include systematics by considering the nuisance parameters α to simply be more parameters we're trying to estimate:

$$\ln L(\theta, \alpha) = \ln L(\theta | D, \alpha) + \ln P(\alpha)$$

The first term is the regular log likelihood---a function of θ , with α considered to be a fixed parameter. The second term is what we call the constraint term---basically it's the prior on α .

Application of constraint terms in likelihood

Remember the problem in which we measured an object using two rulers with different temperature dependencies?

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

$$c_1 = 0.1$$

$$L_1 = 2.0 \pm 0.1$$

$$y_1 = 1.80 \pm 0.22$$

$$T_0 = 25$$

$$c_2 = 0.2$$

$$L_2 = 2.3 \pm 0.1$$

$$y_2 = 1.90 \pm 0.41$$

$$T = 23 \pm 2$$

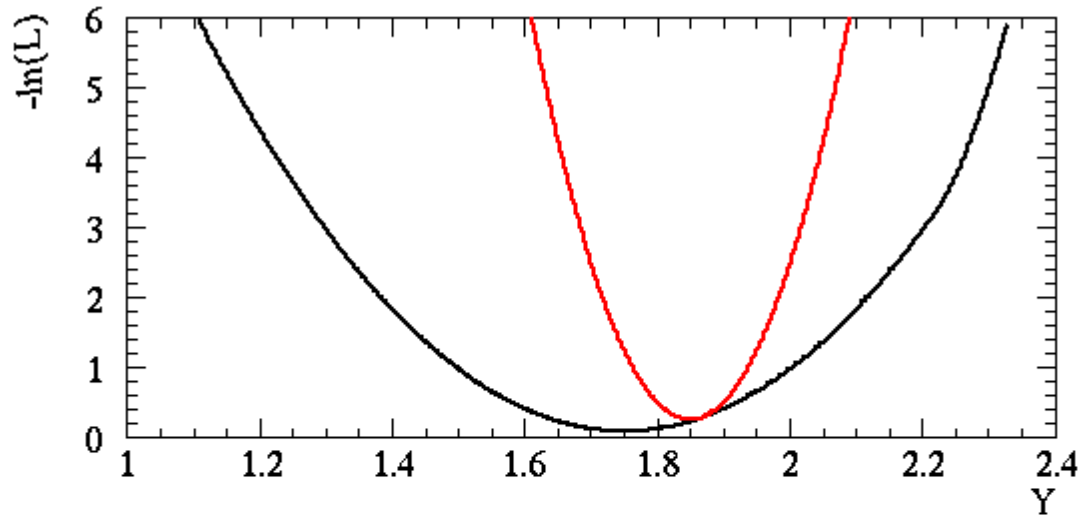
$$\ln L(\theta, \alpha) = \ln L(\theta | D, \alpha) + \ln P(\alpha)$$

$$-\ln L(y, T) = \frac{1}{2} \sum_{i=1}^2 \left(\frac{y - L_i - c_i(T - T_0)}{\sigma_L} \right)^2 + \frac{1}{2} \left(\frac{T - 23}{2} \right)^2$$

The first term of the likelihood is the usual likelihood containing “statistical errors” on the L_i , with T considered fixed. The second is the constraint term (think: “prior on T ”). The joint likelihood is a function of the two unknowns y and T .

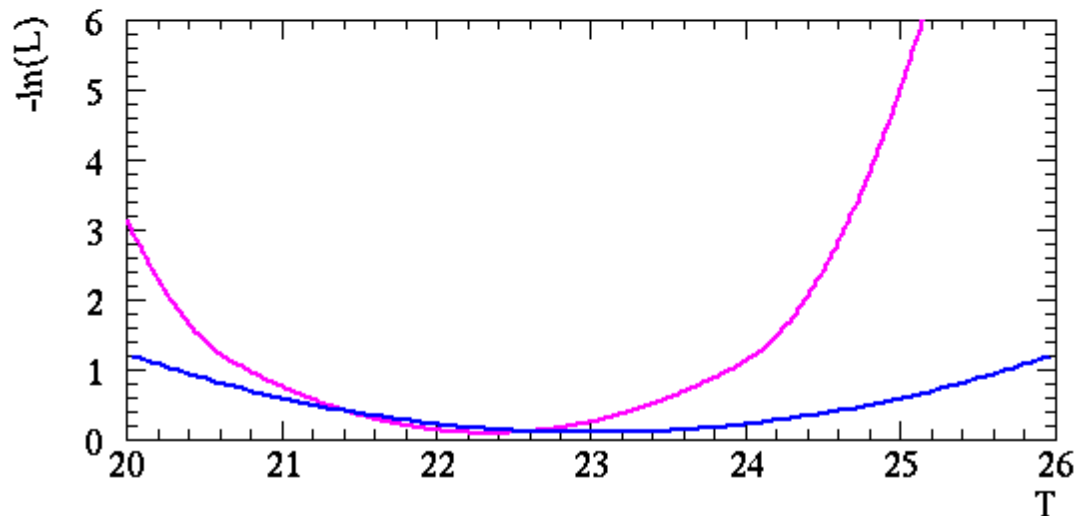
Procedure: marginalize over T to get shape of likelihood as function of y .

Constraint terms in likelihood: results



Top plot is shape of likelihood as function of y , after marginalizing over T :

Red: T fixed (stat error only)
Black: after minimizing $-\ln(L)$ as function of T at each y
 1σ range: same as covariance matrix approach



Blue: "a priori" constraint on T (23 ± 2).
Magenta: shape of likelihood as a function of T , after marginalizing over y .

Gaussian vs. non-Gaussian systematics

The “constraint term” approach to systematics can be used for any systematic---even when the expected distribution of the nuisance parameter is not Gaussian. Consider the joint likelihood:

$$\ln L(\theta, \alpha) = \ln L(\theta|D, \alpha) + \ln P(\alpha)$$

Simply plug in whatever you think the correct form of $P(\alpha)$ to be.

How to report systematics

In reality there is no deep fundamental distinction between statistical and systematic errors. (Bayesians will say that both equally reflect our uncertainty about the universe.) Nonetheless, it is traditional, and useful, to separately quote the errors, such as $X = 5.2 \pm 2.4(\text{stat}) \pm 1.5(\text{sys})$.

There is a common tendency to assume that statistical and systematic uncertainties will be uncorrelated. This is often the case, but not always. (For example, if the data itself is providing a meaningful constraint on the nuisance parameter, there will likely be a correlation.) If such a correlation exists, report it explicitly (maybe as contour plots of X vs. the nuisance parameters). Otherwise you can be sure that someone is going to take your data, add the errors in quadrature, and report

$$X = 5.2 \pm \sqrt{2.4^2 + 1.5^2} = 5.2 \pm 2.8$$

Consider making the full form of the joint likelihood (or the priors and posterior PDFs if it's a Bayesian analysis) publicly available---on the web, if it won't fit in the paper itself.

Systematic uncertainty tables

		X	Y
Energy scale	1.00%	+0.05/-0.04	-0.06/+0.04
Energy resolution	2.00%	+0.03/-0.03	-0.03/+0.03
Cross-section uncertainty	0.50%	+/-0.02	+/-0.02
Detection efficiency	0.60%	+0.03/-0.03	+0.025/-0.025
Fiducial volume	3.00%	+0.15/-0.14	+0.15/-0.14
Total		+0.17/-0.14	+0.16/-0.15

A sample table showing the individual sizes of various systematic uncertainties. The second column is the size of the systematic itself, while the third and fourth columns show the effect of that systematic on two measured quantities X and Y. The order of the signs indicates the correlation between the systematic effects---for example, energy scale here moves X and Y in opposite directions.