C. Blocker
March 22, 2001
Version 1.1

Systematic Uncertainties

Abstract

A brief discussion of systematic uncertainties is given, including how systematic uncertainties are defined, various types of systematic uncertainties, and how systematic uncertainties are determined in some example cases.

1 Introduction

In any measurement, properly determining the uncertainty on the measurement is as important as determining the measured value. This uncertainty is traditionally divided into a statistical part and a systematic part. This note briefly discusses some issues related to the systematic part of the uncertainty, often called the systematic error.

The first issue is to define what is meant by a systematic uncertainty. This is not trivial and is discussed in section 2. Section 3 discusses various types of systematic uncertainties that are commonly encountered. Section 4 covers how to determine systematic uncertainties in some sample cases and how to include them in the uncertainty. Section 5 discusses how to handle potential unknown sources of systematic uncertainties. In section 6, issues regarding systematic uncertainties when combining measurements are discussed. Finally, section 7 gives the conclusions and summary.

2 Definition

The division of the uncertainty between statistical and systematic is clear in some cases, but it is not clear, nor universally agreed, in other cases. To understand some of the issues here, consider a few simple cases.

First, consider the measurement of a cross section for producing a particle by observing a specific decay mode of the particle. For this example, assume that the efficiency ϵ and the integrated luminosity L are well known somehow. The number of events N expected to be found is given by

$$N = \epsilon \sigma B L, \tag{1}$$

where σ is the cross section and B is the branching ratio of the particle to the observed mode. This can be converted to an equation for σ:

$$\sigma = \frac{N}{\epsilon BL}. \tag{2}$$

Thus, if we measure the number of events, we can determine the cross section. If we repeat the experiment with the same integrated luminosity, we will likely observe a different number of events, since the observed number of events is a Poisson distribution random variable. This uncertainty, that is, the difference between the observed number of events and the average number of events that would be observed in a very large number of repetitions of the experiment, is known as the statistical uncertainty or statistical error, since it is due to statistical fluctuations in the observable that determines the quantity of interest.

Next, consider the branching ratio in this example. Branching ratios are measured quantities and also have uncertainties (although in some case these may be very small). Thus, the measured cross section has a contribution to its uncertainty that comes from the uncertainty in the branching ratio. In this example, assume that the branching ratio comes from the Particle Data Book. Then, repeating the experiment would have no effect on the branching ratio. Everyone would agree that in this case the uncertainty on the branching ratio is a systematic error since it has nothing to do with statistical fluctuations in the current experiment.

Given simple considerations such as these, some people have proposed that the statistical uncertainty include any contribution that scales as the square root of the number of events, that is, as the square root of the integrated luminosity, and that the systematic uncertainty include all other uncertainties. This is a possible definition, but it is not what is commonly done.

Consider a possible uncertainty in the efficiency. Efficiencies may be determined in a number of ways, some of which include factors that are determined from the data. These factors will most likely scale as the square root of the number of events. However, most people would include any uncertainty in the efficiency as a systematic uncertainty. This is perhaps because uncertainty in the efficiency usually includes other factors, such as Monte Carlo statistics and model dependence.

Most measurements in experimental particle physics depend (in some form or another) on counting events. A possible working definition is that the statistical error is the uncertainty due to statistical fluctuations of the parameter that goes directly into the measurement, and that the systematic error is the remainder of the uncertainty. There are undoubtedly counterexamples to this, such as, a maximum likelihood fit, where this proposed separation may be difficult in practice.

Thus, the exact separation between statistical and systematic uncertainty is not well defined, but is subject to common practice. If it is not clear in a given case, discussion and common sense will usually lead to a reasonable solution.

3 Types of Systematic Uncertainties

Systematic uncertainties may be classified as one of several types. These types are not unique but may be helpful in thinking about how to handle systematic uncertainties. Most measured values depend on a quantity that is directly measured (often a number of events) and various parameters (such as efficiencies, backgrounds, integrated luminosities, branching ratios, etc.) that correct and convert this directly measured quantity to the desired quantity.

Sometimes these parameters are determined in the same experiment. For example, efficiencies and integrated luminosities are always like this. If the parameter is determined in a separate measurement, even though it is within the same experiment, its uncertainty is usually included as a systematic error. If the parameter is determined simultaneously, as in a maximum likelihood fit, then usually no effort is made to separate the parameter's uncertainty, although this is discussed in more detail below.

Some parameters are taken from other experiments, for example, a branching ratio measured elsewhere might be needed. These are clearly systematic uncertainties.

Sometimes parameters are taken from theory, although this is usually not desirable. Theoretical uncertainty on a parameter is a contribution to the systematic uncertainty of the measurement. Some theoretical parameters have several predicted values or a range of possible values. If there is no guidance from theorists about which values are more reasonable, then taking the root-mean-square variation (assuming a flat distribution in the case of a range of values) is reasonable, although it is also common for people to use an uncertainty that spans all the possible values.

Theory can also enter as models that are used to determined certain parameters. For example, efficiencies are often determined from a Monte Carlo program that has theoretical models such as differential cross sections and parton distributions. If the theoretical model is wrong, then there is an error in the efficiency. Most people determine the systematic uncertainty in this case by varying parameters in the theoretical model and seeing how much the calculated efficiency changes. This raises several questions. First is how much to vary the theoretical parameter. This is usually somewhat a matter of judgement and tradition. For example, it is common to vary the Q^2 parameter in the parton distributions by a factor of two up and down. This is not unique but is conventional. The next question, is how to convert the change in efficiency to a systematic error. Again, there is no unique answer. If only one or a few parameters is varied, then the systematic uncertainty is usually taken to cover all the calculated values (sometimes in an asymmetric manner). If many parameters are varied, then often people add the uncertainties in quadrature.

When several experiments are measuring the same quantity and wish to be able to easily compare or combine results, they will agree on what value, range, and uncertainty to use for some theoretical parameters, as was done by several of the LEP Working Groups. This is convenient but shouldn't prevent people from continuing to think about the best value and uncertainty for a parameter.

When varying theoretical model parameters, it is important to make sure that the model is consistent with the data. For example, if the efficiency is sensitive to the transverse momentum spectrum of the events, then you should check that the theoretical model, including detector effects, reproduces the observed P_t spectrum of the data. This should remain true for all allowed variation of parameters.

For all types of systematic uncertainties, it is important to accurately determine them and be clear in a paper or note about how they were determined. For parameters that were determined elsewhere (either from theory or in another experiment) it is also vital to be clear about what uncertainty was used, how it entered into the total uncertainty, and the sign of the effect. Then, if the measurement or theory of the parameter improves, the published result can be easily updated.

Sometimes people refer to "conservative uncertainties" that are intended to cover any possible variation in a parameter, usually one from theory. This tends to inflate the uncertinty, which sometimes makes people fell more comfortable. However, this also reduces the power and significance of the measurement. Our goal should be get both the correct measurement and the correct, minimal uncertainty. In some cases, this may mean that we have accept a large uncertainty in a theoretical parameter because it really is poorly known. As a counterexample, consider initial state radiaion in the top mass measurment. Originally, people used the range from the Pythia default to no initial state radiation as a "conservative" estimate. It was then correctly pointed out that we know enough about QCD that the default Pythia setting is much more likely than no radiation. The systematic uncertainty was adjusted to a Gassian distribution about the default Pythia result.

4 Determination and Combination of Systematic Uncertainties

There are several ways to determine systematic uncertainties and then to fold them into a measurement. One of the most straight forward is when the measured parameter is given by an equation, such as equation 2, and the systematic uncertainties (such as, those on the branching ratio, efficiency, and integrated luminosity) are determined externally. In that case, the common practice today is to combine the systematic uncertainties using the standard propagation of errors formula, that is, if A is a measured parameter that is a function of a set of variables α_i's with systematic uncertainties σ_{α_i}, then

$$\sigma_A^2 = \sum_i \left(\frac{\partial A}{\partial \alpha_i}\right)^2 \sigma_{\alpha_i}^2. \tag{3}$$

When a measured parameter is determined from a likelihood fit, then inclusion of systematic uncertainties becomes more difficult. Sometimes, this must be done by folding the systematic uncertainty with the fundamental probability distribution. This is discussed in a CDF note by Luc Demortier [1]

Another case that arises in likelihood fits is where other parameters must be simultaneously fit to obtain the measured parameter. For example, considered an unbinned likelihood fit to Gaussian distribution data where the mean is the desired parameter (such as, a fit the the mass of a bump). In this case, the number of signal events, the experiment resolution, and parameters describing the background are all likely to be additional fit parameters. Due to correlations, the uncertainty on the mean will be larger in this case than if all the additional parameters were accurately known. Since untangling these additional contributions is difficult, if not impossible, for a single case, most people just use the error on the measured parameter (the error on the mean in this case) returned from the fit (assuming the fit and error determination are done properly) as the statistical uncertainty, even though it includes contributions from uncertainties in the other parameters.

Another case that arises in likelihood fits is dependence on a parameter that is determined externally (for example, a branching ratio from the Particle Data Book). There are several methods that are used. First, some people do the fit with the parameter fixed to its external value to determine the measured parameter and its statistical error. Then, they vary the externally determined parameter by plus and minus its uncertainty and redo the fit. The problem with this method is that the fit value of the measured parameter may change due to statistical fluctuations as well as systematic dependence on the external parameter. Since the parameter is determined externally, its contribution to the systematic uncertainty shouldn't change as the number of fit events is increased.

This effect can be explored with a simple Monte Carlo program, where you might find that the variation of the fit value of the measured parameter may well decrease as the number events increases until it asymptotically approaches a value, which is the true contribution to the systematic uncertainty (assuming the Monte Carlo reasonably models the process). Another equivalent method is to generate a large number of pseudo-experiments, each with the same number of events as the data sample. As a parameter is varied, any variation of the average measurement of these pseudo-experiments is a systematic uncertainty. The effect of the statistical uncertainty (or any other uncertainty) contributing more than once is known as double counting and is to be avoided. Note in this case that if there are a large number of such external parameters, the cumulative double counting can be significant.

Another way that people incorporate externally determined parameters into a likelihood fit to include a constraint, namely, a factor of the form (assuming the uncertainty on the parameter is Gaussian)

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}, \tag{4}$$

where x is the parameter in the fit (free to float), x_0 is its externally determined value, and σ is the uncertainty in that determination. In this case, if a correlation exists, the uncertainty on the externally determined parameter will contribute to the uncertainty from the fit on the measured parameters. Attempts to disentangle this contribution will probably lead to confusion and the possibility of double counting.

If the data have the power to determine an externally known parameter, then a third method for handling this would be to just allow the parameter to be free in the fit with no constraint. Any systematic uncertainty is then automatically included in the uncertainty on the measured parameter. In addition, you would get an new, independent measurement of the externally known parameter.

Which of these methods to use for including an externally known parameter in the likelihood function depends somewhat on the circumstances. If the parameter is known much more accurately than the power of the data to determine it, then fixing it to externally measured value is the right thing to do. For example, if the likelihood function depended on the mass of the electron, no one would let it float, either constrained or unconstrained. In this case, the value is so well known that in most cases, people wouldn't even worry about a possible contribution to the systematic uncertainty. On the other hand, if the data have much more power to determine the parameter than the external uncertainty, letting the parameter float in the fit and getting a new measurement of it is sensible. For the many cases in between, the constrained fit is probably the right answer.

It should be noted as a general principle that inclusion or increase of systematic uncertainties should always increase the total uncertainty and increase upper limits. If you find the opposite behaviour, then you should be suspicious of a mistake or a faulty algorithm.

5 Unknown Sources of Systematic Uncertainties

There is always the possibility of a systematic uncertainty that did not occur to you. One method used to check this is to change analysis cuts, histogram binning, fit method, etc. and study changes in the measurement. Due to statistical effects (for example, changing a selection criterion will change the events in the sample), changes are expected, and these changes should not be added tot the systematic uncertainty, since to do so is another form of double counting.

However, if the changes are outside what is significantly outside what is expected, you should look carefully for an additional source of systematic uncertainty. If one if found, you may be able to at least partially correct for it, and you then include the residual systematic uncertainty. If no source of the variation is found, then you must decide whether the changes are due to a larger than expected statistical fluctuation or whether you need to add an additional systematic uncertainty.

An example of this occurred in the $\sin 2\beta$ analysis where the dilution for the same side tagging depended more sensitively on the p_T cut on the tracks. See reference [2] for more details.

6 Combining Measurements with Systematic Uncertainties

When combining measurements of the same parameter, there are a couple of issues regarding systematic uncertainties that should be considered. First, it is possible to have common systematic uncertainties, even when the measurements are from different experiments. For example, if two experiments measure the same cross section by using the same value for a relevant branching ratio, then any error in that branching ratio will affect each experiment and their average equally. Thus, when combining multiple measurements, it is important to identify common systematic errors, remove them before the statistical combination of the measurements, and then add them back to the new result. This, of course, can get complicated when the dependence on the common parameter is not as clear and/or the correlation is not 100%, in which case an error matrix approach is appropriate.

Another issue that arises is the case where the measurements don't agree. For example, if there are two measurements of a parameter that differ by ten standard deviations, the most likely scenario is that one or both of the experiments got the measurement wrong (that is, this is usually considered to be much more likely than the difference being due to statistical fluctuations). If it is not known which experiment is wrong, then usually a weighted average is calculated and an additional systematic uncertainty is included so that both measurements are covered.

A good example of combining measurements with correlated systematic uncertainties is the combined CDF and D0 measurement of the top mass (reference [3]).

7 Conclusions and Summary

It is difficult to define systematic uncertainty in a way that conforms with all common practice. This fact and the nature of systematic uncertainties, makes general statements and numeric analysis difficult. Several instances of common practice as understood by the author have been presented here. The general conclusion is that the experimenter must think carefully about systematic uncertainty, so that when taken together the statistical and systematic uncertainty accurately reflect the quality of the measurement.

References

[1] Luc Demortier, “Folding Systematic Uncertainties into Probability Density Functions”, CDF Note 5305.

[2] Note on systematic errors in the $\sin 2\beta$ analysis is being prepared.

[3] Note on systematic errors in the top mass analysis is being prepared.

A Uncertainties on Uncertainties

Suppose a parameter α is determined to be $\alpha_M = \alpha_0 \pm \sigma_\alpha$, where there is also an uncertainty on σ_α. that is, $\sigma_\alpha = a \pm b$. The question becomes what is the correct uncertainty to use for α.

To determine this, consider the probability distribution for α, which is a convolution of a Gaussian for α with a Gaussian for σ_α, that is,

$$P(\alpha) = \int_{-\infty}^{\infty} \frac{d\sigma}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\alpha-\alpha_0)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi b^2}} e^{-\frac{(\sigma-a)^2}{2b^2}}. \tag{5}$$

So far, I have found this integral to be intractable. However, the rms standard deviation of α can be calculated, which is really what we want for the systematic uncertainty. First, the average value of α is

$$\overline{\alpha} = \int_{-\infty}^{\infty} \alpha P(\alpha) d\alpha = \alpha_0, \tag{6}$$

where the trick is to do the α integration first. The average value of α^2 is

$$\overline{\alpha^2} = \int_{-\infty}^{\infty} \alpha^2 P(\alpha) d\alpha = \alpha_0^2 + a^2 + b^2. \tag{7}$$

Thus, the uncertainty on α is

$$\Delta\alpha = \sqrt{\overline{\alpha^2} - \overline{\alpha}^2} = \sqrt{a^2 + b^2}. \tag{8}$$

Thus, if b is significant when added in quadrature with a, then it should be included. However, since σ_α must be positive, if b is sizeable compared to a, the probability distribution for σ_α cannot be Gaussian. One way to handle this is to make the lower limit of the integral in equation 5 zero. Another option is to put the actual probability distribution for σ_α into equation 5, if it is known. In either case, the integrals will probably have to be done numerically.