What is systematics ?

Giovanni Punzi

February 9, 2001

Abstract

I propose here a clear–cut and general definition of what is a systematic uncertainty, with the aim of helping making clear distintions and avoiding double–counting of errors.

1 Introduction

Systematic uncertainty is a very familiar concept for all experimental physicists; yet if you ask exactly what it is, you get an array of different answers, most of them rather vague. It is generally assumed that everybody can understand for himself what it is after having seen a few examples. This mostly works, but there are indeed cases where it is not so clear what should be counted as systematics and there is even a risk of double–counting some uncertainties. In those cases, it is the lack of a clear definition that makes it hard to be absolutely sure that one is doing the "right thing".

To help sorting out this confusion, especially inside my own head, I tried to find a clear definition that would decide on ambiguities, and be a good representation of the existing feeling on the subject. To expain the idea, I will take a step back, and start from what the *statistical* uncertainty is.

2 A rigorous definition of systematics

When we make a measurement, we try to learn about some unknown parameter μ by observing the value of some other (observable) quantity x, that is somehow related to μ (both of them may be multidimensional). If x is just a fixed function of μ: $x = x(\mu)$, than there is no uncertainty in inferring the value of μ, except possibly for some mathematical indeterminacy if the function is not invertible.

In real life, however, the relationship between x and μ is always of a probabilistic nature, that is, x is a random variable with some probability distribution that depends on the value of μ: $p(x;\mu)$. Even if the same value of μ can yield different values of x, we can still infer something about μ, but with some uncertainty, that we call *statistical uncertainty*. Its size is determined by the shape of the function $p(x;\mu)$: the wider the larger. This function contains all information about the physics processes involved and the apparatus used in the observation, and in deriving the statistical uncertainty it is assumed to be a perfectly definite, unchanging, and known function. In real life, even this ("second–order") knowledge cannot be absolutely perfect; actually, it is likely to be rather approximate when complex phenomena and apparati are involved. Not only you don't know exactly what value of x will be observed if μ has some definite value, but you don't even know exactly how probable each possible value of x is. This further uncertainty will reflect in an additional uncertainty on the inferred value of μ, which I propose to call *systematic uncertainty*. To summarize the above in a single phrase:

> "Systematic uncertainty is, in any statistical inference procedure, the uncertainty due to incomplete knowledge of the probability distribution of the observables"

There are several advantages in adopting the above definition for deciding what must be counted as systematics, which I will try to outline in the following. It appears clearly that this definition is simple and rigorous and can be applied to any problem. I will try to show that it gives the "intuitive answer" in all obvious cases, prevents double–counting, and provides a more solid justification for many customary practices. In other words, it really helps you to calculate systematics and to get a clearer vision of many practical problems.

3 Notes

3.1 Simple examples

The simplest examples are where the pdf is perfectly known in functional form, but contains an unknown parameter. Let's look at measuring the cross section of a particle[1]. In this case you do know that the distribution of your counts

[1] first Craig's example

is going to be a Poisson, but the mean mu depends on a constant, L, which is only approximately known. So this is clearly a case of systematic uncertainty, and you must properly keep it into acount in quoting the final result.

Second, suppose you measure something with a gaussian error, but with a instrument that has a calibration uncertainty. This effectively means that you don't know exactly the average of your gaussian, so again this is an uncertainty in the pdf, and falls in the systematic category.

It should be also apparent that all such things as a Q^2 scale, and structure functions fall in the systematics category.

3.2 Quoting systematics

One other thing that appears clearly from the proposed definition of systematics is that there is a clear separation between what is statistics and what is systematics: they really belong to two different species. This justifies the fact that systematics errors get a different treatment, and are generally calculated and quoted separately from other errors. It is also clear that often the best representation of systematic uncertainty is by a "range" rather then by a sigma. In fact, the degree of knowledge of a systematic parameter cannot be represented by a distribution, otherwise you would be able to incorporate it in the global pdf, and it will not be a systematic anymore (but see below on Bayesian tretment). In a way, a systematic uncertainty can never be exactly defined.

3.3 Extra parameters

Often the uncertainty in a pdf comes from the dependence on a certain number of parameters whose value is partially or totally unknown. The *functional form* of the pdf might be exactly known, but it may depend on some physical quantities of uncertain value (e.g. temperature, Q^2 scale, luminosity, structure functions....), so that the pdf as a whole is *uncertain*. In other cases, one may want to artificially introduce additional arbitrary parameters, just as a way to parametrize a space of functions spanning the extent of our uncertainty in the space of the possible pdf's, which on turn may have no explicit dependence on any uncertain physical parameter.

Sometimes, the real pdf of our measurement may actually depend on some physical parameter without us being aware of it, and without our knowledge of the functional form or the physical basis for this dependence. This situation may go under the name of "systematic effects of unknown origin".

In fact, it is generally considered good experimental practice to repeat a measurement under different external conditions, even if they do not appear explicitly in the expression of the pdf that one is assuming. If a significant variation is found, that means that there is an unanticipated *systematic effect*. If one is unable to compensate for it in some way, than one needs to add an uncertainty to account for it. In cases where the variation is really unexpected, however, one probably wants to investigate the matter a little further, rather then simply add some extra contribution to the systematic uncertainty, as the real meaning of what happened is that our model of reality has proved wrong, so we might be making a much larger error that we think.

3.4 Ill-defined parameters

In some cases, the effect of the systematic uncertainty may be to obscure the very meaning of the parameter begin measured. For instance, if we want to measure the slope of a straight line, and the curve is a parabula instead, one may quote a systematic, but there is the important question of establishing just *what* is being measured as “slope” in this case. For instance, the situation is very different if the experimental procedure actually measures the “average slope” of the curve, or if it measures, for instance, the slope at the origin. Depending on that, it is even possible that the uncertainty on the actual shape of the line produces zero systematic uncertainty (remember that the systematic uncertainty is about the parameter, not the pdf in itself.)

3.5 Bayes and systematics

A peculiar situation arises if you have a Bayesian attitude. In this case, you may elect to convert all systematic uncertainty in a statistical one. In fact, after parametrizing the uncertainty in the pdf by a certaing number of parameters, you can assume an a-priori distribution for each of them, and calculate a resulting pdf by integrating over all those parameters. This resulting ”smeared” pdf will be then exactly known, and therefore the systematic will be zero, at the price of an increased statistical error. Clearly one has then a result that depends on the priors; in some lucky situations the parameters might even have well-defined distributions because of the physics involved, such that even the strictest frequentist will be willing to average over them; in many other cases the prior choice will contain some element of subjectivity.

It is important to note that it is not allowed to use a likelihood function in place of a prior for the purpose of this averaging; a likelihood is not a probability distribution, and to trasform it into one you need to multiply it by some other (previously existing) prior, so you can't escape the issue of choosing it.

In any case, however, it is important to realize that integrating over systematic parameters may easily lead to non-indipendent inferences in different measurements ("correleated errors"). That is, if I make two different measurements that depend on the uncertain value of the luminosity L, and in both cases get rid of it by integrating over some a-priori distribution of L, then the two final quoted statistical uncertainties will become partially correlated. This may have a great influence, for instance, if a third quantity is calculated that depends on both.

3.6 Systematics in fits

The question of what to do in fits becomes clear, in the light of this definition.

- If the systematic parameters are taken as additional free parameters in the fit, there is no uncertainty in the pdf, so there is no systematic. The error is purely statistical (and rightly so). This puts on a solid ground what is a common practice[2].
- When you fix the parameter in the fit function from an external source, then a systematic appears, of course, related to how well it is known. The range of variation of the output results, obtained by exploring the full range allowed of the parameter, give the error. If the parameter comes from an independent measurement, a possible way to transform this in a range might be to calculate a Confidence Region at some appropriate CL.
- Constrained fit: also in this case, there is no systematic. The stat uncertainty you find is smaller than in case 1), but you pay this with the inclusion of the "other" measurement inside your dataset, that means, your statistical error will now be correlated with that of whatever other measurement uses the same constraint.

3.7 double counting examples

.

With a clear definition, you avoid double counting. For instance, it is not right to vary the bin widths in a fit to an histograms, and to add the variation of the fit parameters as an additional systematics. It should be clear that in

[2]see Craig's , bottom of page 4)

this case there is no uncertainy in the pdf (if the fit function is perfectly known, the pdf is given just by the Poisson distribution of the bin contents), so from the proposed definition, there is no systematic. The changes in the paramenters one observes is in reality a double counting of the statistical uncertainty. In fact, binning data reduces information, and two different binnings of the same sample amount to different choices of the pieces of information to discard. The uncertainty associated with this reduction of information content is already kept into account in the larger uncertainty returned by a binned fit with respect to an unbinned fit: no extra contributions need to be added. With the above definition, it is immediately clear that there is no systematic involved here.

The same can be said, e.g. of varying the position of the cuts used to select the sample, or the range where the fit is performed. All those things have nothing to do with uncertainties in the pdf, so they should not be counted as systematics. Clearly, if a large variation is observed (larger that the calculated statistical uncertainty) you should suspect that there is something wrong, as this amount to a failed goodness-of-fit test. For instance, if the fit results change significantly by choosing a different fit range, this means the shape of the data is different from the shape of the fitting function. When this happens, the right thing to do is not to add a systematic contribution and go on with the measurement, but rather to review critically the choice of the fit function and its uncertainty. Anyway, it seems sensible to do this kind of checks via an explicit, standard goodness–of–fit test instead.

Despite the above, it might still be a good idea to quote the effect of some of this variations, but for a different purpose, that is to convince the reader that you didn't make an ad–hoc choice based on data to force a particular result. For instance, if the significance of a peak attains its maximum just for the particular choice of cuts adopted for the analysis, it is hard to convince anybody that the cuts were not tuned based on the sample at hand. It is important just to make sure that this changes are not added to the uncertainty as if they were systematic contributions.