



## Systematics and Limit Calculations

Wade Fisher  
*Fermilab*  
(Dated: December 28, 2006)

This note discusses the estimation of systematic uncertainties and their incorporation into upper limit calculations. Two different approaches to reducing systematics and their degrading impact on upper limits are introduced. An improved  $\chi^2$  function is defined which is useful in comparing Poisson distributed data with models marginalized by systematic uncertainties. Also, a technique using profile likelihoods is introduced which provides a means of constraining the degrading impact of systematic uncertainties on limit calculations.

### I. INTRODUCTION

A typical search analysis is ultimately described by a final variable (or variables) chosen to be sensitive to a parameter of the search system. The result of the search are distributions of this final variable for the signal process, one or more modeled background processes, and the observation from data. These final variable distributions become the input to a calculation of upper limits on the signal model search parameter[1].

The signal and background model predictions generally depend on a range of ingredients which are necessary for the measurement, but are not of immediate interest. Examples of such parameters are integrated luminosities, efficiencies, acceptances, and calculated cross sections. Referred to as “nuisance parameters”, the values of these ingredients are important in the extraction of limits on the parameter of interest, and any uncertainty in these nuisance parameters

will necessarily degrade sensitivity to the parameter of interest.

This note describes two techniques useful for reducing the impact of systematic uncertainties on limit calculations. The utility of these techniques are certainly not limited to limit calculations, although this will be the context in which the ideas will be introduced.

## II. CONSTRAINTS ON SYSTEMATIC UNCERTAINTIES

Systematic uncertainties are, in general, estimated rather than measured (Sec.A2). Thus, the values estimated frequently correspond to an upper bound of the true value. This overestimation, although appropriately cautious, necessarily leads to the degradation of calculated limits. However, the magnitude of systematic uncertainties can be constrained by comparing the predicted shape and rate of the final variable to what is observed in data. Physicists routinely rely on the degree of agreement in the final variable to search for indications of improper analysis construction. To do so, they will calculate a  $\chi^2$  for the final variable distribution, relying on the difference in observed and predicted values, as well as the uncertainties on the predicted values:

$$\frac{\chi^2}{ndof} = \frac{1}{ndof} \sum_i \left( \frac{p_i - d_i}{\sigma_i} \right)^2 \quad (1)$$

where  $p_i$  is the predicted value,  $d_i$  is the observed data,  $\sigma_i$  is the uncertainty on the prediction, and  $ndof$  is the number of degrees of freedom for the sample (generally taken to be the number of bins less one)[8]. The rule of thumb is that values near unity indicate both good modeling and proper estimation of systematic uncertainties; values below(above) unity indicate overestimated(underestimated) systematic uncertainties. This treatment is usually as much as this useful tool is exploited. Fortunately, with a bit of reformulation, this tool can be expanded to provide constraints on the size of systematic uncertainties via the statistical power inherent to the shape and rate of the final variable distribution.

### A. A More General $\chi^2$

The discussion in this subsection follows largely from [2]. An improved  $\chi^2$  function has been proposed in statistical discussions [3, 4]. The general argument begins with the relationship of the common  $\chi^2$  function to the likelihood function in the limit of Gaussian statistics where, ignoring systematics:

$$\chi^2 = \sum_i \frac{(p_i - d_i)^2}{\sigma_i^2} \simeq -2 \ln(\mathcal{L}/\mathcal{L}_0) \quad (2)$$

$$\mathcal{L} = \prod_i \left( \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp^{-\frac{(p_i - d_i)^2}{2\sigma_i^2}} \quad (3)$$

$$\mathcal{L}_0 = \prod_i \left( \frac{1}{\sigma_i \sqrt{2\pi}} \right) \quad (4)$$

where  $\mathcal{L}_0$  corresponds to a perfect fit,  $p_i = d_i$ . This function, unfortunately, fails to describe properly all cases in which Gaussian statistics does not apply. However, in these cases it is straight forward to reintroduce the Poisson statistical likelihood function:

$$\mathcal{L} = \prod_i \frac{p_i^{d_i} \exp^{-p_i}}{d_i!} \quad (5)$$

$$\mathcal{L}_0 = \prod_i \frac{d_i^{d_i} \exp^{-d_i}}{d_i!} \quad (6)$$

Using this new definition, we can write the associated  $\chi^2$  function as:

$$\chi^2 = 2 \sum_i \left[ (p_i - d_i) - d_i \ln \left( \frac{p_i}{d_i} \right) \right] \quad (7)$$

The impact of systematic uncertainties can be introduced by a transformation of the predicted values:

$$p'_i = p_i \prod_{k=1}^K (1 + f_k^i S_k) \quad (8)$$

where the  $p'_i$  are the systematically varied prediction in bin  $i$  and the values  $f_k^i$  describe the size of each of the  $K$  nuisance parameters. The values  $S_k$  represent Gaussian constraints with zero mean and unit width.

### B. Constraining Uncertainties

Given this improved  $\chi^2$  function, we can now move toward an interpretation of the impact of systematic uncertainties. We have constructed a  $\chi^2$  function that is sensitive to fluctuations in the per-bin predictions. In fact, the response of the  $\chi^2$  function to the nuisance parameters provides a measure of the true variance associated with that nuisance parameter[5]. Because the condition for minimizing the  $\chi^2$  function is that the first partial derivative with respect to each parameter ( $a_j$ ) vanish, we can expect that near a local minimum in any parameter, the  $\chi^2$  function can be described with a quadratic function of that parameter. Thus, the uncertainty of the parameter  $a$  in terms of the curvature of the  $\chi^2$  function in the region of the minimum is given by:

$$\sigma_j^2 = 2 \left( \frac{\partial^2 \chi^2}{\partial a_j^2} \right)^{-1} \quad (9)$$

It is worth noting that when using this improved  $\chi^2$  function, the covariance matrix associated with the Gaussian statistical treatment does not exist. Because equation 9 is derived assuming a curvature matrix, we must be somewhat careful in interpreting this result. In particular, it is unsafe to assume the response of the  $\chi^2$  function will be symmetric. Thus, when extracting a variance using this approach, the positive and negative deviations should be calculated independently.

Using this prescription, one may utilize the inherent information stored in the shape and rate of the observed final variable to place an upper limit on the value of systematic uncertainties. This technique amounts to providing an extended utility of the common technique of calculating the  $\chi^2$  for a final variable.

## III. REDUCING THE IMPACT OF SYSTEMATICS

As outlined in [6], the choice of test statistic which is used in a  $CL_s$  confidence level calculation should be made in an effort to optimize the separation of the signal (signal plus background,  $H_S$ ) and null (background-only,  $H_0$ ) hypotheses. In general, an optimal solution is the likelihood ratio:

$$Q = \frac{P(\text{data}|H_S)}{P(\text{data}|H_0)} \quad (10)$$

As has been noted above, the probabilities  $P(\text{data}|H_S)$  and  $P(\text{data}|H_0)$  are inherently sensitive to how well the models are specified. That is, systematic uncertainties have the effect of broadening the probability distribution functions (PDFs) and thereby reducing all instantaneous probabilities and degrading the separation power of the likelihood ratio.

For a given set of predictions, observations, and systematic uncertainties, it is common for an experiment to find the “best-fit” model for the data. In such cases, the null hypothesis is adjusted such that the PDFs are maximized over the space of all possible values of the systematic uncertainties. This procedure is generally referred to as the “profile likelihood” technique. With this PDF maximization in mind, the optimal choice of test statistic can be redefined. The technique of profiling thus has the benefit of reducing the degree to which the two model PDFs ( $P(\text{data}|H_S)$  and  $P(\text{data}|H_0)$ ) are smeared by the process of marginalization (*i.e.*, fluctuation of predicted values as in equation 8). For a more complete discussion of profiling and marginalization, the reader is referred to [7].

## A. The Best-Fit Model

Given a set of predictions, observations, and systematic uncertainties, one can define a model which represents the best fit to the data observation within the constraints of the systematic uncertainties. A reliable means of performing this fit is achieved by minimizing the  $\chi^2$  function in equation 7. The input to the problem is the *bestestimates* for each background source model and systematic uncertainty. To properly incorporate these best estimates, we must modify equation 7:

$$\chi^2 = 2 \sum_i \left[ (p'_i - d_i) - d_i \ln \left( \frac{p'_i}{d_i} \right) \right] + \sum_k S_k^2 \quad (11)$$

where the definitions of  $p'_i$  and  $S_k$  are adopted from equation 8. In this way, the  $S_k$  can be fit to minimize equation 11.

To properly incorporate the profile likelihood approach into a limit calculation, the profile likelihood must be recalculated each time a pseudoexperiment is simulated. This is equivalent to performing a best fit for each simulated outcome of the experiment.

## B. Application of the Profile Likelihood

The prescription for obtaining the profile likelihood outlined above is not, in general, complicated. However, the application of this technique must be performed with care and several caveats exist. The author will admit that making this technique fully generalizable is difficult, but a consistent treatment can be obtained with the following considerations.

### 1. Signal Contamination

Although the issue of signal contamination can be generalized, there are instances in which specialized solutions are advantageous. The general case should be discussed first.

To accommodate a signal of unspecified size and distribution, the profile likelihood must be generated separately for the signal and null hypotheses, for each pseudoexperiment. In this way, the statistical treatment retains the full dichotomy of hypotheses, *i.e.*, each hypothesis is treated as true independently. Depending on the size and distribution of the signal model, the resulting PDFs may vary significantly. This solution has the advantage of properly maintaining any correlations in systematics present between signal and background models.

In cases in which the signal model is concentrated in a few bins of the final variable distribution, a side-band fit technique can be used. In this case, one must choose a signal/background ratio cut-off above which bins will be excluded from the fit. This solution has the advantages of requiring only one fit and rendering the fit partially insensitive to the shape and size of the signal distribution.

### 2. Grouping of Uncertainties

To speed up computation time, uncertainties can be grouped according to their nature. For example, given several systematics that each represent uncertainties about the overall rate (*e.g.*, luminosity), they can be grouped to one value to be fit. However, this introduces an implicit correlation between the uncertainties and the uncorrelated values cannot be easily extracted. In many instances this distinction is small, but care must be taken to ensure a consistent treatment.

### 3. Distribution Binning

Clearly, the degree to which the shape of the final variable distribution can be used to constrain uncertainties is determined by the binning of the distribution. To maximize the ability to constrain and reduce the noise of the system, the binning should be chosen to reflect the experimental resolution of the final variable. By utilizing a binning that is finer than the experimental resolution, any fit will be more susceptible to random fluctuations and noise, thereby masking true sensitivity.

## IV. CONCLUSIONS

In this note, we have discussed two means of treating systematic uncertainties and their relationship to limit calculations. These two methods are intended to place upper limits on the size of uncertainties and to reduce the overall impact on limit calculations via profiling. These two techniques are intended to be generally useful for many applications involving systematic uncertainties.

### APPENDIX A: IMPACT AND ESTIMATION OF SYSTEMATICS

#### 1. Impact of Uncertainties on Limit Calculations

As noted in [1], systematic uncertainties can be easily incorporated into a limit calculation by smearing the Poisson probability distribution functions (PDFs) for the signal and background expectations. This treatment has the effect of broadening the PDFs and reducing the implicit separation between the two hypotheses being tested: the background-only and signal plus background hypotheses. The amount an upper limit will degrade under the influence of systematic uncertainties depends primarily on the size of the parent distributions being modeled. As the limits are taken from Poisson-distributed predictions of repeated experiments, the Poisson uncertainty associated with the central value is always present. The size of the systematic uncertainties relative to this baseline width will determine their impact. Thus, large systematics become increasingly important with increasing number of observed events.

There is a finite probability that Gaussian fluctuations of nuisance parameters will fall to negative values. This behavior is unphysical and cannot be accommodated by the statistical technique described in [1]. The common solution is to truncate these Gaussians at or very near zero. This treatment is generally unobtrusive until uncertainties reach roughly 30% of the nominal predicted value. Beyond this, any limit extracted will carry a bias from the truncation. It is most desirable to discover a means to avoid this unwanted behavior.

#### 2. Estimating Systematic Uncertainties

To say that systematic uncertainties are measured is rarely accurate. In truth, most systematic uncertainties are only estimated by approximation. For example, the efficiency for selecting electrons with a given set of quality requirements depends on whether the electron-selection algorithm is applied to data or Monte Carlo (MC) events. To correct for this difference, a scaling factor can be supplied to convert between the two. This correction for a measurement that is systematically mismeasured cannot be known to infinite precision, as it depends on a very large number of varying factors: the uniformity of calorimeter calibrations, uniformity of tracking efficiency, accuracy of MC modeling, etc. Thus, we introduce an uncertainty on the systematic correction factor that is large enough to accommodate our relative lack of precise knowledge. This systematic uncertainty is thus an educated estimation of an upper limit.

In very many instances, the true impact of this systematic uncertainty on an analysis is further estimated by propagating the value through the analysis selection. Generally, the nominal value for a systematic correction is varied by  $\pm 1\sigma_{corr}$  to evaluate the  $\pm 1\sigma$  impact for the analysis. This technique often has the unfortunate drawback of convoluting a systematic uncertainty with a binomial probability associated with a restriction (or cut) on that variable. For example, uncertainties on the correction factor for the electron energy scale will move certain events in and out of an analysis selection depending on the minimum electron energy requirement. Thus, the impact to the analysis now depends upon the precise shape of the distribution from which the correction factor is drawn.

The uncertainty associated with a systematic correction generally depends on how it was evaluated. Many systematic corrections and their associated uncertainties are measured directly in data events. One drawback of this approach arises when the statistical uncertainty on the correction being measured is the largest portion of the total systematic uncertainty. However, systematic corrections to theoretical cross sections, for example, cannot generally be compared to data and the uncertainty is related to the precision of the calculation. The degree to which such calculations are compatible with data observations is then not factored into the value of the uncertainty.

These comments are not intended to suggest systematic uncertainties are poorly measured in general. However, to assume any precision or accuracy for systematic uncertainties is rarely safe. It is worth noting the limitations of such

estimates and their common upper-limit nature.

---

- [1] Wade Fisher, “Calculating Limits for Combined Analyses”, DØ Note #4975.
- [2] T. Junk, “Building a More General  $\chi^2$ ”, CDF/DOC/STATISTICS/PUBLIC/7904.
- [3] S. Baker and R. D. Cousins, “Clarification of the Use of Chi-Square and Likelihood Functions in Fits to Histograms”, Nucl. Instrum. Meth. **A221** 437 (1984).
- [4] T. Devlin, “Correlations from Systematic Corrections to Poisson-Distributed Data in Log-Likelihood Functions”, CDF/DOC/JET/PUBLIC/3126.
- [5] P. R. Bevington and D. K. Robinson, “Data Reduction and Error Analysis for the Physical Sciences”, 2nd Ed., ISBN 0-07-911243-9.
- [6] A. L. Read, J. Phys. G **28**, 435 (2002).  
T. Junk, Nucl. Instrum. Meth. A **434**, 435 (1999).
- [7] T. Junk, “Sensitivity, Exclusion, and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties”, CDF/DOC/STATISTICS/PUBLIC/8128.
- [8] In alternative formulations, the  $\sigma_i$  values can be defined as the square root of the data value in bin  $i$ .