Incorporation of Systematic Uncertainties on Search Results

Tom Junk
FNAL Conf. Level Workshop
March 2000

- Deciding on an Approach
- Simple Examples
- A full example -- LEP Higgs combination

Also known as: "Nuisance Parameters"

Taking Systematic Uncertainties into Account

Searches for new particles generally have:

- Small numbers of expected signal and background events
- Signal and background constrained to be not less than zero
- Both signal and background estimates are uncertain. Only the data counts are known with certainty.

The often-used approach:

Described in
Cousins and Highland, Nucl. Instr. Meth. **A320**, 331 (1992).

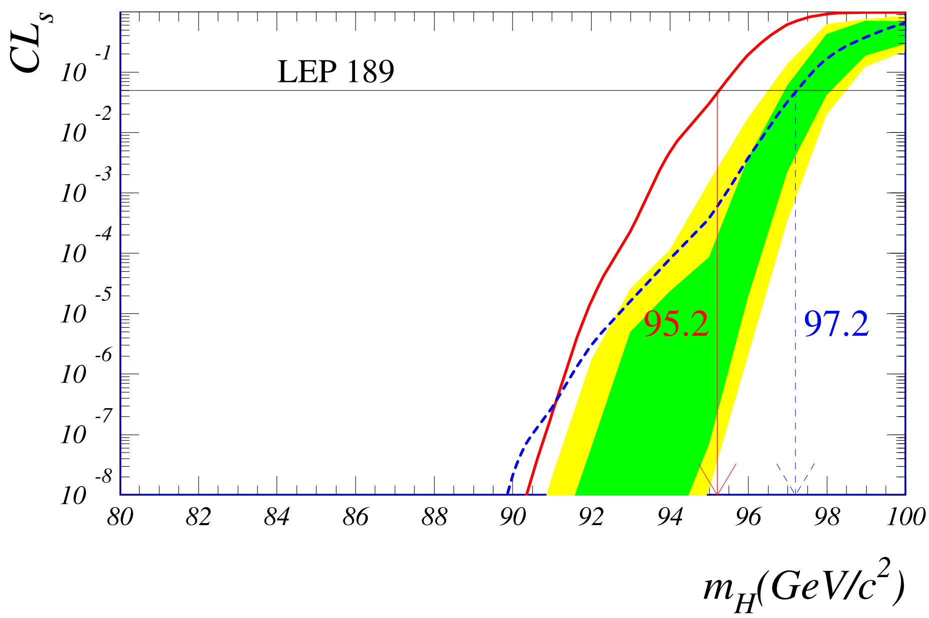
- Prescription: average the CL over an ensemble of values for the uncertain parameters: ***Bayesian!***
- C+H: analytic, signal errors only, single seach channel. But easy to extend.

(e.g., Giunti, hep-ex/9901015 for bg err)

Averaging a Confidence Level

- Already a problem -- averaging a confidence level over an ensemble of possible experimental outcomes can produce a very unlikely expected CL due to a tail to high CL's.

Comb. of LEP 189 GeV data (when we used the average expected CL)



- An improvement -- use the median conf. level when quoting a sensitivity expectation

- *But medians are incorrect for syst. errors.* Simple example: single counting experiment, single source of uncertainty. The median CL is the one you get with no error treatment at all!

Features of Average Conf. Levels

- Upper bounds depend quadratically on syst. errors if they are small -- "intuitive"
- Effect of syst. errors using C+H's prescription is usually small, relative to the now very old prescription: subtract 1σ from the background and signal when forming limits. *Add* 1σ to the background when deciding on a discovery. Conservative! (and "flip-floppy")

- Sometimes one can exclude more model hypotheses when larger systematic errors are included.
  - Only if one has candidates with small expected background with large errors.
  - The sensitivity (expected limit) worsens with the inclusion of systematic uncertainty, the actual limits may get more stringent!
- And -- naive application of Average CL's to high-statistics two-sided limits is not appropriate! (can even get empty confidence intervals, averaging in many possible exclusions and only a few acceptances if the stat. error << syst. error)

No single "unified" treatment of syst. errors in the low-stats 1-sided error case and the high-stats 2-sided limit case.

Bayesian or Classical?

- Ideally, all nuisance parameters should be constrained by measurements
  - Need to trace all uncertainties to their original sources -- external measurements

- Producers of results have full access to the raw data and Monte Carlo -- they should produce results independent of priors.
- Consumers of results usually have no choice - they are given a confidence interval for a measured quantity and cannot form P(data|hypothesis) for another experiment but may be able to approximate it.
- In incorporating systematic uncertainties one plays the role of consumer *and* producer.

- A mixture of Classical and Bayesian techniques is therefore almost unavoidable.

A Completely Classical Approach

If one has access to

- P(data|hypothesis) **and**
- the ensemble of possible experimental outcomes

For the measurements of each nuisance parameter,

THEN

One can make a joint Confidence Interval calculation using

P(all data|all hypothesized parameters)
and the joint ensemble.

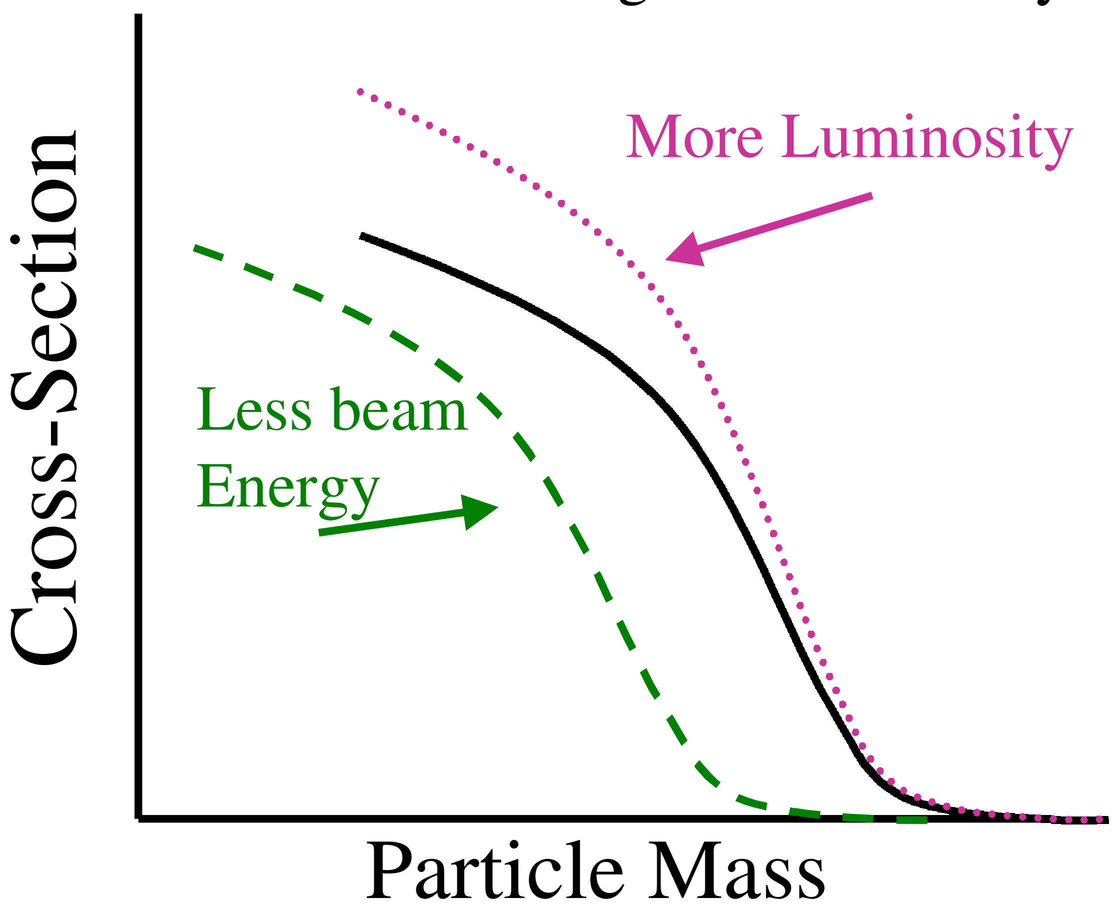
Very Difficult! (lots of work)
And you have to decide on a common model
And P(something's mismeasured somewhere)≈1

Similar in spirit to the LEP+SLD Electroweak SM fits (high-statististics with Gaussian Errors).
Interesting side effect:
A measurement's sensitivity to a nuisance parameter may provide a better "measurement" of the nuisance parameter than the intended parameter.

Not All Nuisance Parameters are Created Equal

Example: Search for a new particle pair-produced in e^+e^- collisions (kinematic threshold)

Nuisance parameters: Accelerator energy
Integrated luminosity



$$\frac{\partial \text{limit}}{\partial \text{beam energy}} = 1$$

Syst. error on limit=syst. error on beam energy

But: better to use C+H for the luminosity error

Systematic Uncertainty on the Model??

Suggestion -- do ***not*** include these in the Confidence Level calculation for the model being tested.

Reason -- models can be tested with data even if they are incorrect.

But experimentalists should

- Fully specify the model which is tested
- Describe in detail the sensitivity of the results to variations in the nuisance parameters of the model or the use of alternative calculations.
- Seek to present results which are as independent of models as possible, in addition to presenting model-dependent results.

But: Systematic uncertainties in models used to unfold effects not intended to be tested by the measurement need to be considered in the CL.

Practical Details

Example -- Single counting experiment, low statistics.

$$CL_{s+b} = \sum_{x \le x_{obs}} p(\text{data} \mid s+b)$$

x="test statistic"

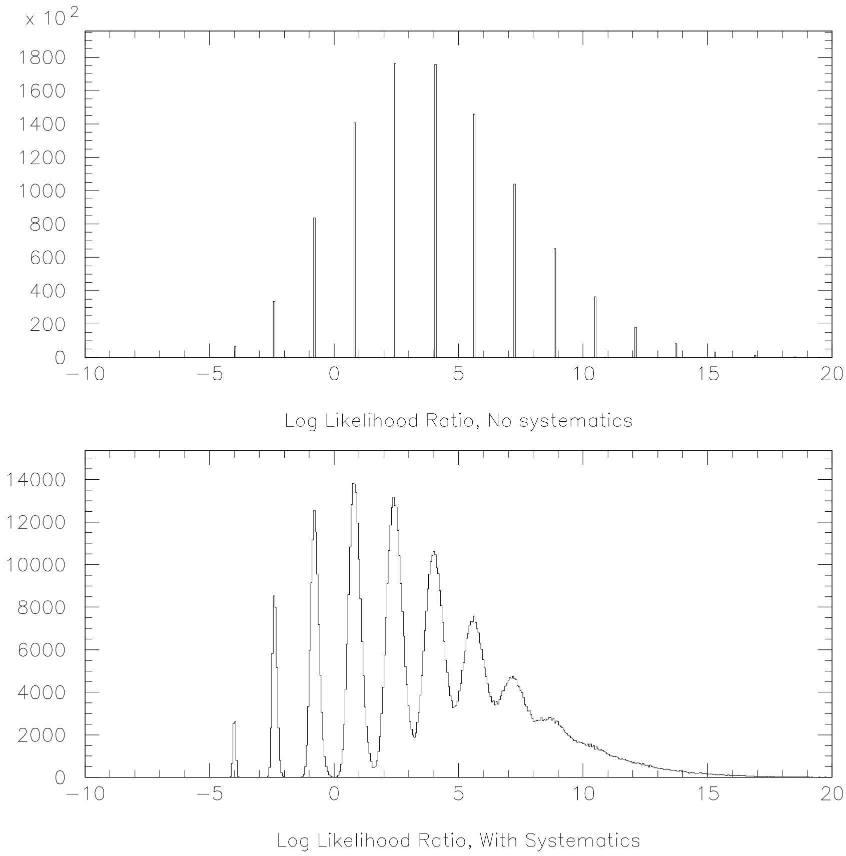
Use the Log Likelihood Ratio to order outcomes (trivial in this example, but necessary for combining unequal counting experiments or bins of the same experiment)

$$x = \log\left(\frac{p(n \mid s+b)}{p(n \mid b)}\right) = -s + n\log(1 + s/b)$$

(Poisson)

Typical calculations use the PDF of x in the $s+b$ and b hypotheses. This is modified by systematic uncertainties on s and b.

Effect of Systematic Errors on the PDF of x



Need to be careful not to "split" outcomes in the sum of the probabilities for the Confidence Level.

Single channel case: No effect on ordering of outcomes so no test-statistic is needed.
The above picture is irrelevant for the single channel case!

Multichannel Case: Two options

Multichannel Combinations With Uncertainties

- Exact procedure still under discussion within the LEP Higgs working group.

The problem:

Both the probabilities of the possible outcomes of the experiment and the ordering of these outcomes (based on the likelihood ratio or other sensible test-statistic) depend on the values of the nuisance parameters.

Two alternatives:

- Hold the ordering fixed but average the outcome probabilities
- Average the CL allowing the ordering to vary with the nuisance parameters.

Both reduce to Cousins+Highland's prescription for a single counting experiment.
Both have analogues in standard procedures.

Varying the Ordering or Not

Case for varying the ordering:

One must evaluate the effects of the nuisance parameters everywhere we use them to interpret the data. We may not know if a particular outcome looks more like a signal than another because of our lack of knowledge of some nuisance parameter.

Case against varying the ordering:

One ought to fix the model at its best-guess value, use it to analyze the data, and then test to see what would happen if it is incorrect -- the nuisance parameters can have different values in Nature than they have in our Monte Carlos.

A selection can be optimized as one sees fit, but to evaluate syst. errors one must fix the selection technique and vary the possible data it might see.

Not settled!

The Full Application -- Combining Higgs Search Results from Four LEP Experiments

- Similar Experimental Apparatus
- Same Accelerator
- Same Monte Carlo Models (choice of a few)
- Search for the same particle(s) in the same decays
- Differing sensitivity/resolution/background/ search techniques
- A mixture of shared and independent error sources
- Time pressure, political importance

To claim evidence for new particle production in a combination of experiments, errors on the background need to be understood with their correlations included. -- 4x the luminosity will magnify any mismodeling and *can fake a signal*.

Exchanging LEP Higgs Results for Combination

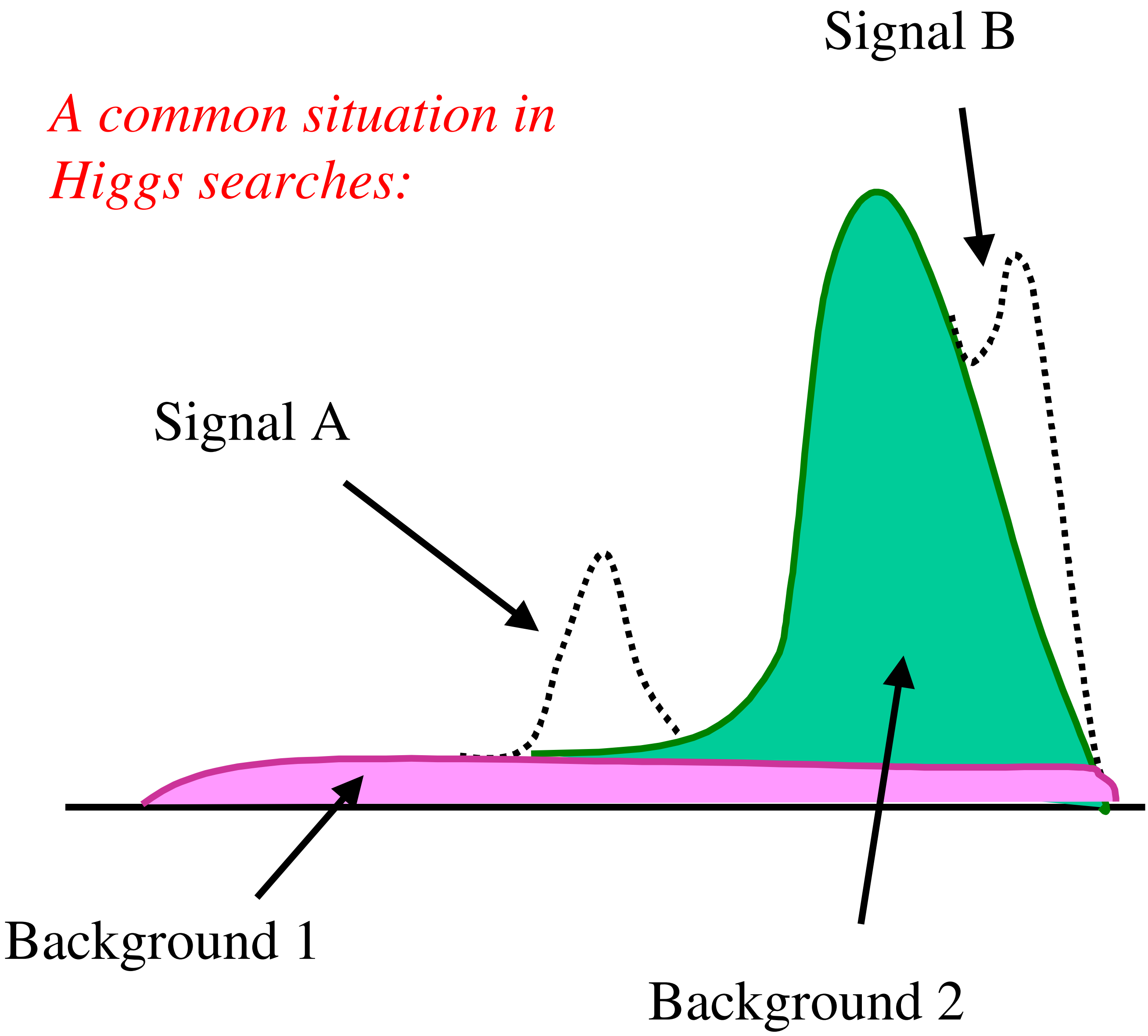
- Results are binned in an experiment's choice of discriminant variables (reconstructed mass, b-tag, kinematic variables, or a combination)
- For each bin:
  - signal estimation
  - background estimation
  - data counts
- For each search analysis channel:
  - **signed** relative error on the signal itemized *by named source*
  - signed relative error on the background itemized by named source
- Example sources (full list is very long):
  - e^+e^- -> ZZ and WW cross-sections
  - B-hadron decay charged multiplicity
  - B-hadron fragmentation function
  - 4-jet rate in e^+e^- -> qqbar events
  - Comparison of standard MC generators
- Common central values and systematic variations used -- guided by precision measurements.

LEP Higgs Error Exchange

- Signal and Background estimates can share common uncertainties and therefore be correlated.
- Errors with the same name are considered 100% correlated (value and sign important!)
- Errors with different names are considered 0% correlated
- Arbitrarily correlated errors can be decomposed into correlated and uncorrelated parts
- In practice, most errors are uncorrelated
  - On OPAL, Monte Carlo statistics often dominates the uncertainty
  - Many different sources of correlated errors, but the sources are uncorrelated, so the effect is to reduce the correlation.
- Prior practice -- consider *all* errors as correlated (very conservative) Properly handling correlations strengthens the limit and the justifiability.

Not Included Yet -- Shape Errors

A common situation in Higgs searches:



Depending on the Higgs mass, the background could come mainly from source #1 or source #2 and have different uncertainty. Reporting errors on each bin is a lot of overhead. Can weight error by its importance, or divide the search into two or more channels.

Effect of Syst. Error Treatment on the LEP SM Higgs Limit

All errors uncorrelated: almost no effect on the sensitivity or actual limits.

All errors 100% correlated with the same signs: about 200-300 MeV reduction of expected and actual SM Higgs mass limit.

Errors properly correlated: Around 100 MeV reduction in expected and actual limits.

Effect is not pronounced for the 95% CL exclusion point, but is a large factor for $CL_s << 1$. It is important to be prepared for a discovery!