Systematic Uncertainties: Principle and Practice

Outline

1. Introduction to Systematic Uncertainties
2. Taxonomy and Case Studies
3. Issues Around Systematics
4. The Statistics of Systematics
5. Summary

Pekka K. Sinervo,F.R.S.C.
Department of Physics
University of Toronto

Introduction

- **Systematic uncertainties play key role in physics measurements**
  - **Few formal definitions exist, much "oral tradition"**
  - **"Know" they are different from statistical uncertainties**

Random Uncertainties

- Arise from stochastic fluctuations
- Uncorrelated with previous measurements
- Well-developed theory
- Examples
  - **measurement resolution**
  - **finite statistics**
  - **random variations in system**

Systematic Uncertainties

- Due to uncertainties in the apparatus or model
- Usually correlated with previous measurements
- Limited theoretical framework
- Examples
  - **calibrations uncertainties**
  - **detector acceptance**
  - **poorly-known theoretical parameters**

Literature Summary

- **Increasing literature on the topic of "systematics"**

  A representative list:

  – R.D.Cousins & V.L. Highland, NIM **A320**, 331 (1992).
  – C. Guinti, Phys. Rev. D **59** (1999), 113009.
  – G. Feldman, "Multiple measurements and parameters in the unified approach," presented at the FNAL workshop on Confidence Limits (Mar 2000).
  – R. J. Barlow, "Systematic Errors, Fact and Fiction," hep-ex/0207026 (Jun 2002), and several other presentations in the Durham conference.
  – G. Zech, "Frequentist and Bayesian Confidence Limits," Eur. Phys. J, **C4:12** (2002).
  – R. J. Barlow, "Asymmetric Systematic Errors," hep-ph/0306138 (June 2003).
  – A. G. Kim et al., "Effects of Systematic Uncertainties on the Determination of Cosmological Parameters," astro-ph/0304509 (April 2003).
  – J. Conrad et al., "Including Systematic Uncertainties in Confidence Interval Construction for Poisson Statistics," Phys. Rev. D **67** (2003), 012002
  – G.C.Hill, "Comment on "Including Systematic Uncertainties in Confidence Interval Construction for Poisson Statistics"," Phys. Rev. D **67** (2003), 118101.
  – G. Punzi, "Including Systematic Uncertainties in Confidence Limits", CDF Note in preparation.

I. Case Study #1: W Boson Cross Section

- **Rate of W boson production**
  - **Count candidates N_s+N_b**
  - **Estimate background N_b & signal efficiency ε**

$$\sigma = (N_c - N_b)/(\varepsilon L)$$

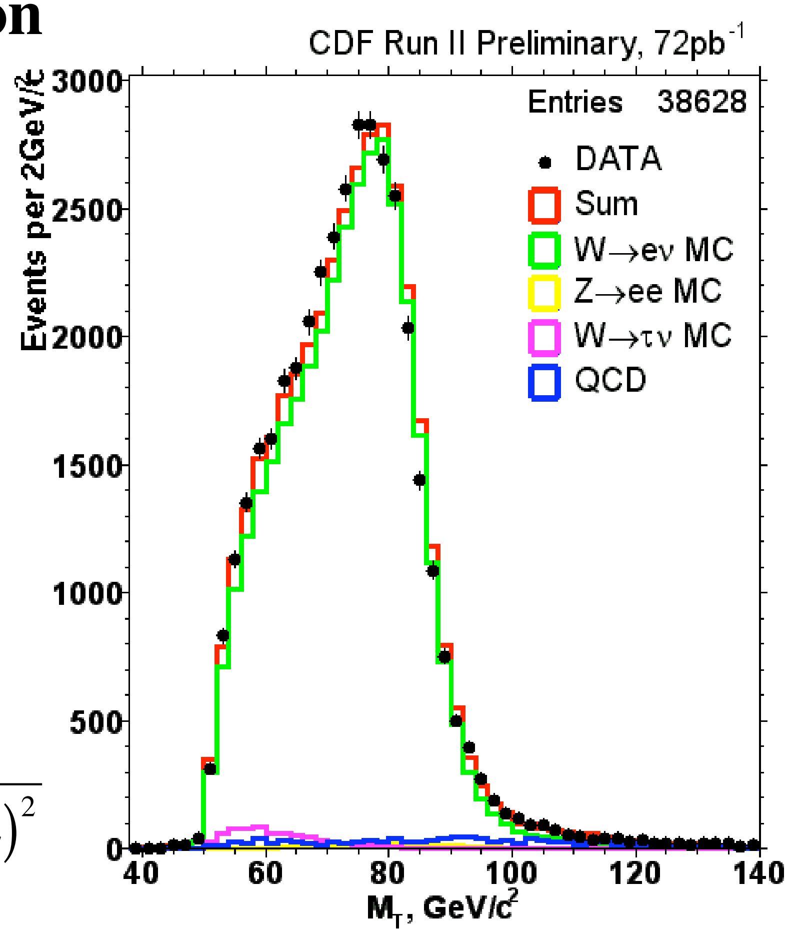
  - **Measurement reported as**

$$\sigma = 2.64 \pm 0.01\,(\text{stat}) \pm 0.18\,(\text{syst})\ \text{nb}$$

  - **Uncertainties are**

$$\delta\sigma_{stat} \approx \sigma_0\sqrt{1/N_c}$$

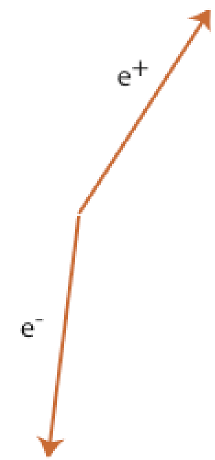
$$\delta\sigma_{syst} \approx \sigma_0\sqrt{(\delta N_b/N_b)^2 + (\delta\varepsilon/\varepsilon)^2 + (\delta L/L)^2}$$



Definitions are Relative

- **Efficiency uncertainty estimated using Z boson decays**
  - **Count up number of Z candidates N_Z^{cand}**
    - Can identify using charged tracks
    - Count up number reconstructed N_Z^{recon}

$$\varepsilon = \frac{N_Z^{recon}}{N_Z^{cand}} \pm \delta\varepsilon \approx \sqrt{\frac{N_Z^{recon}\left(N_Z^{cand} - N_Z^{recon}\right)}{N_Z^{cand}}}$$



  - **Redefine uncertainties**

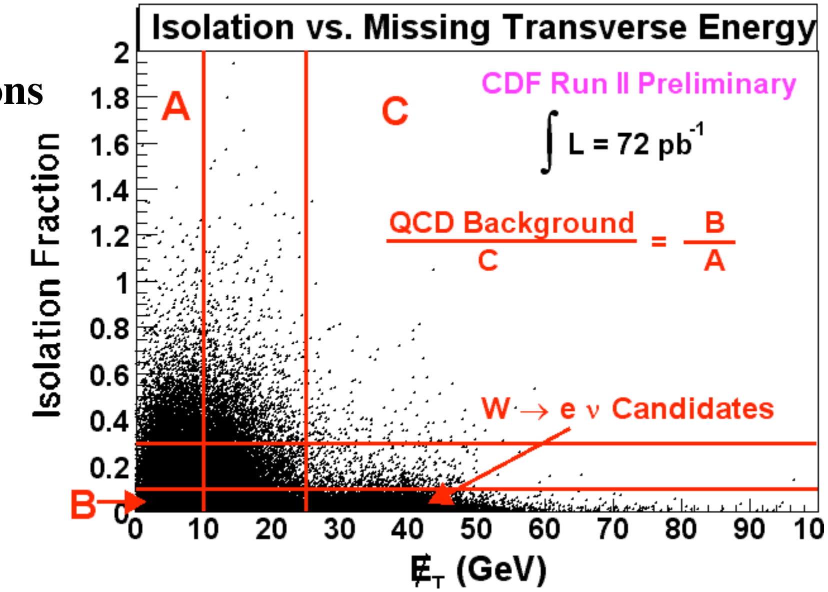
$$\sigma_{stat} \approx \sigma_0\sqrt{1/N_c + (\delta\varepsilon/\varepsilon)^2}$$

$$\sigma_{syst} \approx \sigma_0\sqrt{(\delta N_b/N_b)^2 + (\delta L/L)^2}$$

Lessons:
- Some systematic uncertainties are really "random"
- Good to know this
  - Uncorrelated
  - Know how they scale
- May wish to redefine
- Call these
  "CLASS 1" Systematics

Case Study #2: Background Uncertainty

- **Look at same W cross section analysis**
  - **Estimate of N_b dominated by QCD backgrounds**
    - Candidate event
      - **Have non-isolated leptons**
      - **Less missing energy**
    - Assume that isolation and MET uncorrelated
    - Have to estimate the uncertainty on N_b^{QCD}



  - **No direct measurement has been made to verify the model**
  - **Estimates using Monte Carlo modelling have large uncertainties**

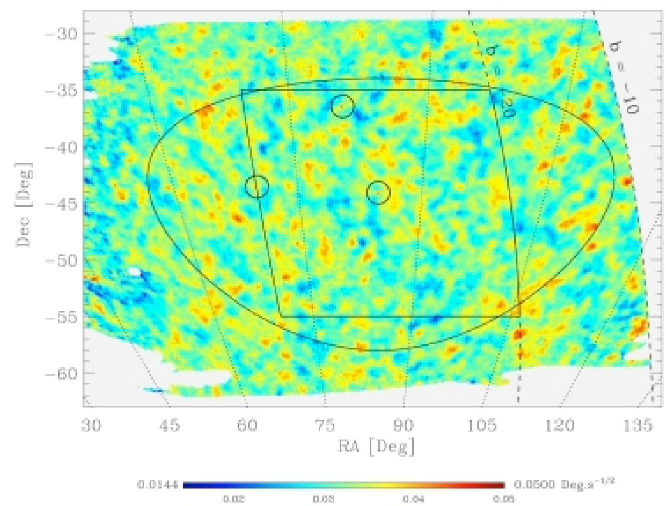
Estimation of Uncertainty

- **Fundamentally different class of uncertainty**
  - **Assumed a model for data interpretation**
  - **Uncertainty in N_b^{QCD} depends on accuracy of model**
  - **Use "informed judgment" to place bounds on one's ignorance**
    - Vary the model assumption to estimate robustness
    - Compare with other methods of estimation
- **Difficult to quantify in consistent manner**
  - **Largest possible variation?**
    - Asymmetric?
  - **Estimate a "1 σ" interval?**
  - **Take** $\sigma \approx \frac{\Delta}{\sqrt{12}}$?

Lessons:
- Some systematic uncertainties reflect ignorance of one's data
- Cannot be constrained by observations
- Call these
  "CLASS 2" Systematics

Case Study #3: Boomerang CMB Analysis

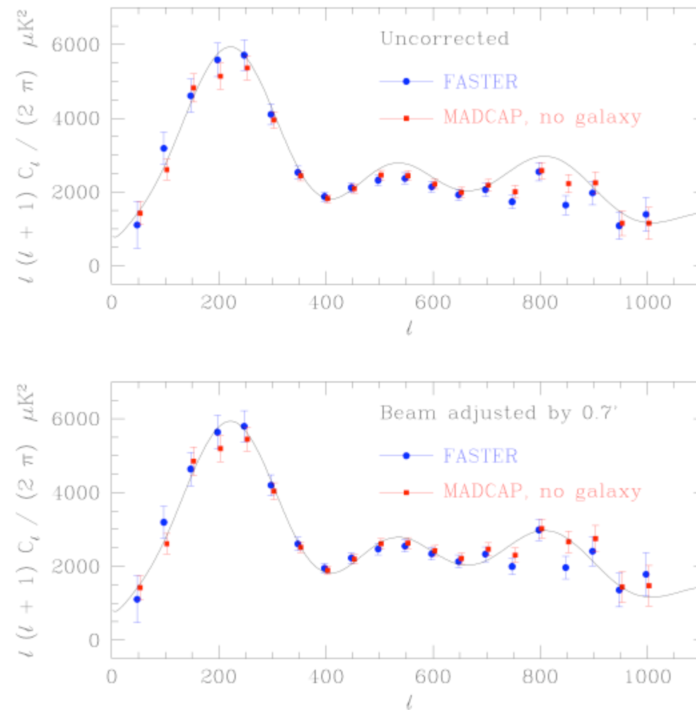
- **Boomerang is one of several CMB probes**
  - **Mapped CMB anisoptropy**
  - **Data constrain models of the early universe**



- **Analysis chain:**
  - **Produce a power spectrum for the CMB spatial anisotropy**
    - Remove instrumental effects through a complex signal processing algorithm
  - **Interpret data in context of many models with unknown parameters**

Incorporation of Model Uncertainties

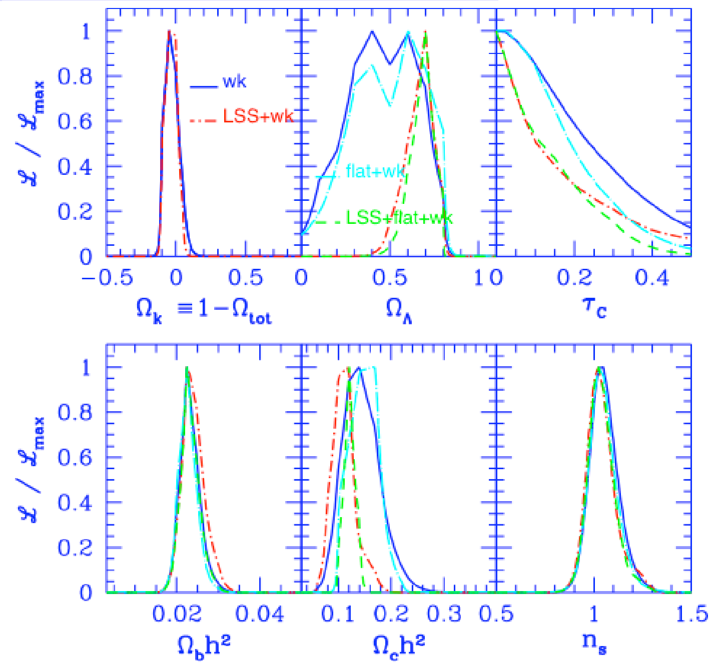
- **Power spectrum extraction includes all instrumental effects**
  - **Effective size of beam**
  - **Variations in data-taking procedures**



- **Use these data to extract 7 cosmological parameters**
  - **Take Bayesian approach**
    - Family of theoretical models defined by 7 parameters
    - Define a 6-D grid (6.4M points), and calculate likelihood function for each

Marginalize Posterior Probabilities

- Perform a Bayesian “averaging” over a grid of parameter values
  - Marginalize w.r.t. the other parameters
    - NB: instrumental uncertainies included in approximate manner
  - Chose various priors in the parameters



- Comments:
  - Purely Bayesian analysis with no frequentist analogue
  - Provides path for inclusion of additional data (eg. WMAP)

Lessons:
- Some systematic uncertainties reflect paradigm uncertainties
- No relevant concept of a frequentist ensemble
- Call these “CLASS 3” Systematics

Proposed Taxonomy for Systematic Uncertainties

- **Three "classes" of systematic uncertainties**
  - **Uncertainties that can be constrained by ancillary measurements**
  - **Uncertainties arising from model assumptions or problems with the data that are poorly understood**
  - **Uncertainties in the underlying models**
- **Estimation of Class 1 uncertainties straightforward**
  - **Class 2 and 3 uncertainties present unique challenges**
  - **In many cases, have nothing to do with statistical uncertainties**
    - Driven by our desire to make inferences from the data using specific models

II. Estimation Techniques

- **No formal guidance on how to define a systematic uncertainty**
  - **Can identify a possible source of uncertainty**
  - **Many different approaches to estimate their magnitude**
    - Determine maximum effect Δ

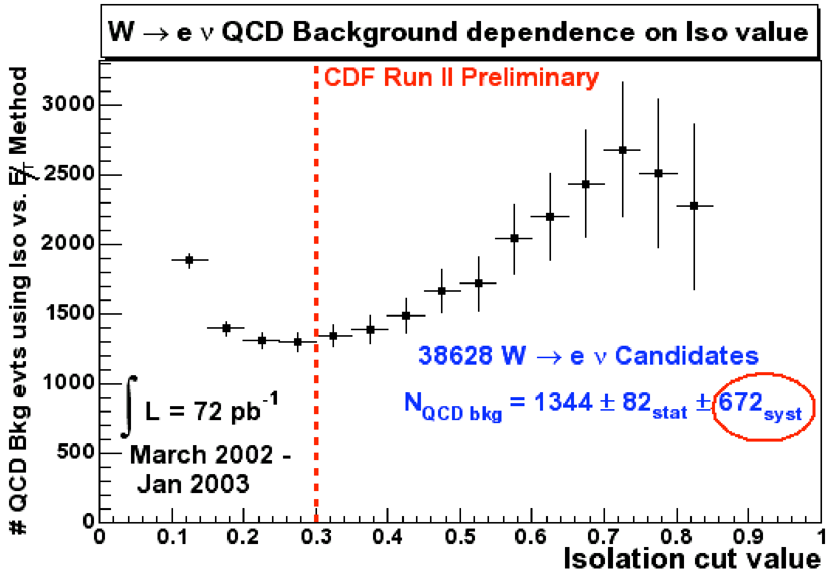
$$\sigma = \frac{\Delta}{2} ?$$

$$\sigma = \frac{\Delta}{\sqrt{12}} ?$$

- **General rule:**
  - **Maintain consistency with definition of statistical intervals**
  - **Field is pretty glued to 68% confidence intervals**
  - **Recommend attempting to reflect that in magnitudes of systematic uncertainties**
  - **Avoid tendency to be "conservative"**

Estimate of Background Uncertainty in Case Study #2

- **Look at correlation of Isolation and MET**
  - **Background estimate increases as isolation "cut" is raised**
  - **Difficult to measure or accurately model**
    - Background comes primarily from very rare jet events with unusual properties
    - Very model-dependent



- **Assume a systematic uncertainty representing the observed variation**
  - **Authors argue this is a "conservative" choice**

Cross-Checks Vs Systematics

- **R. Barlow makes the point in Durham**
  - **A cross-check for robustness is not an invitation to introduce a systematic uncertainty**
    - Most cross-checks confirm that interval or limit is robust,
      - **They are usually not designed to measure a systematic uncertainty**
- **More generally, a systematic uncertainty should**
  - **Be based on a hypothesis or model with clearly stated assumptions**
  - **Be estimated using a well-defined methodology**
  - **Be introduced *a posteriori* only when all else has failed**

III. Statistics of Systematic Uncertainties

- **Goal has been to incorporate systematic uncertainties into measurements in coherent manner**
  - **Increasing awareness of need for consistent practice**
    - Frequentists: interval estimation increasingly sophisticated
      - **Neyman construction, ordering strategies, coverage properties**
    - Bayesians: understanding of priors and use of posteriors
      - **Objective vs subjective approaches, marginalization/conditioning**
  - **Systematic uncertainties threaten to dominate as precision and sensitivity of experiments increase**
- **There are a number of approaches widely used**
  - **Summarize and give a few examples**
  - **Place it in context of traditional statistical concepts**

Formal Statement of the Problem

- **Have a set of observations x_i, $i=1,n$**
  - **Associated probability distribution function (pdf) and likelihood function** $\mathrm{p}(x_i \mid \square) \,\square\, \mathcal{L}(\square) = \square_i \mathrm{p}(x_i \mid \square)$
    - Depends on unknown random parameter $\square$
    - Have some additional uncertainty in pdf
      - **Introduce a second unknown parameter $\square$**

$$\mathcal{L}(\square,\square) = \square_i \mathrm{p}(x_i \mid \square,\square)$$

- **In some cases, one can identify statistic y_j that provides information about $\square$**

$$\mathcal{L}(\square,\square) = \square_{i,j} \mathrm{p}(x_i, y_j \mid \square,\square)$$

  - **Can treat $\square$ as a "nuisance parameter"**

Bayesian Approach

- **Identify a prior □(□) for the "nuisance parameter" □**
  - **Typically, parametrize as either a Gaussian pdf or a flat distribution within a range ("tophat")**
  - **Can then define Bayesian posterior**

$$\mathcal{L}(\square,\square)\ \square(\square)\ d\square\ d\square$$

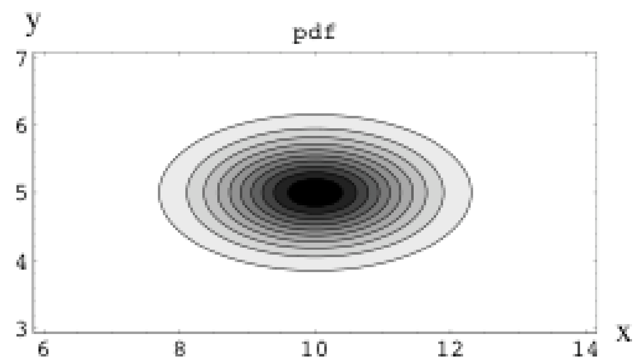
  - **Can marginalize over possible values of □**
    - Use marginalized posterior to set Bayesian credibility intervals, estimate parameters, etc.
- **Theoretically straightforward ….**
  - **Issues come down to choice of priors for both □, □**
    - No widely-adopted single choice
    - Results have to be reported and compared carefully to ensure consistent treatment

Frequentist Approach

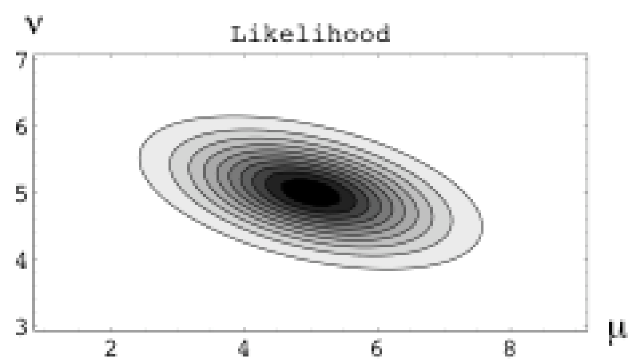
- **Start with a pdf for data** $p(x_i, y_j \mid \mu, \nu)$
  - **In principle, this would describe frequency distributions of data in multi-dimensional space**
  - **Challenge is take account of nuisance parameter**
  - **Consider a toy model**

$$p(x, y \mid \mu, \nu) = G(x - (\mu + \nu), 1)\, G(y - \nu, s)$$

> Parameter s is Gaussian width for ν



- **Likelihood function (*x=10*, *y=5*)**
  - **Shows the correlation**
  - **Effect of unknown ν**



Formal Methods to Eliminate Nuisance Parameters

- **Number of formal methods exist to eliminate nuisance parameters**
  - **Of limited applicability given the restrictions**
  - **Our "toy example" is one such case**
    - Replace x with $t=x-y$ and parameter ν with

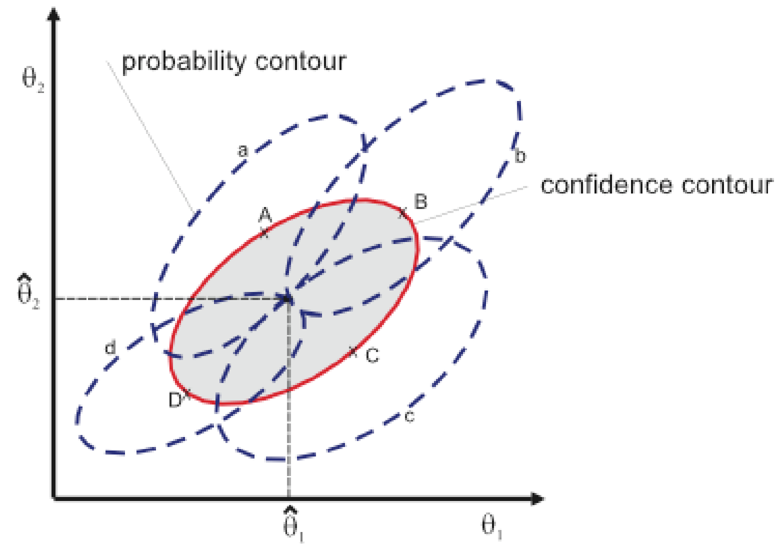
$$\nu' \equiv \nu + \mu s^2 \big/ 1+s^2$$

$$\Rightarrow \; p(t,y \mid \mu,\nu') = G\left(t-\mu, \sqrt{1+s^2}\right) G\left(y-\nu' + ts^2 \big/ 1+s^2,\; s \big/ \sqrt{1+s^2}\right)$$

    - Factorized pdf and can now integrate over ν'
    - Note that pdf for μ has larger width, as expected
  - **In practice, one often loses information using this technique**

Alternative Techniques for Treating Nuisance Parameters

- **Project Neyman volumes onto parameter of interest**
  - **"Conservative interval"**
  - **Typically over-covers, possibly badly**
- **Choose best estimate of nuisance parameter**
  - **Known as "profile method"**
  - **Coverage properties require definition of ensemble**
  - **Can possible under-cover when parameters strongly correlated**
    - Feldman-Cousins intervals tend to over-cover slightly (private communication)



From G. Zech

Example: Solar Neutrino Global Analysis

- **Many experiments have measured solar neutrino flux**
  - **Gallex, SuperKamiokande, SNO, Homestake, SAGE, etc.**
  - **Standard Solar Model (SSM) describes ν spectrum**
  - **Numerous "global analyses" that synthesize these**
- **Fogli et al. have detailed one such analysis**
  - **81 observables from these experiments**
  - **Characterize systematic uncertainties through 31 parameters**
    - 12 describing SSM spectrum
    - 11 (SK) and 7 (SNO) systematic uncertainties
- **Perform a χ^2 analysis**
  - **Look at χ^2 to set limits on parameters**

Hep-ph/0206162, 18 Jun 2002

Formulation of χ^2

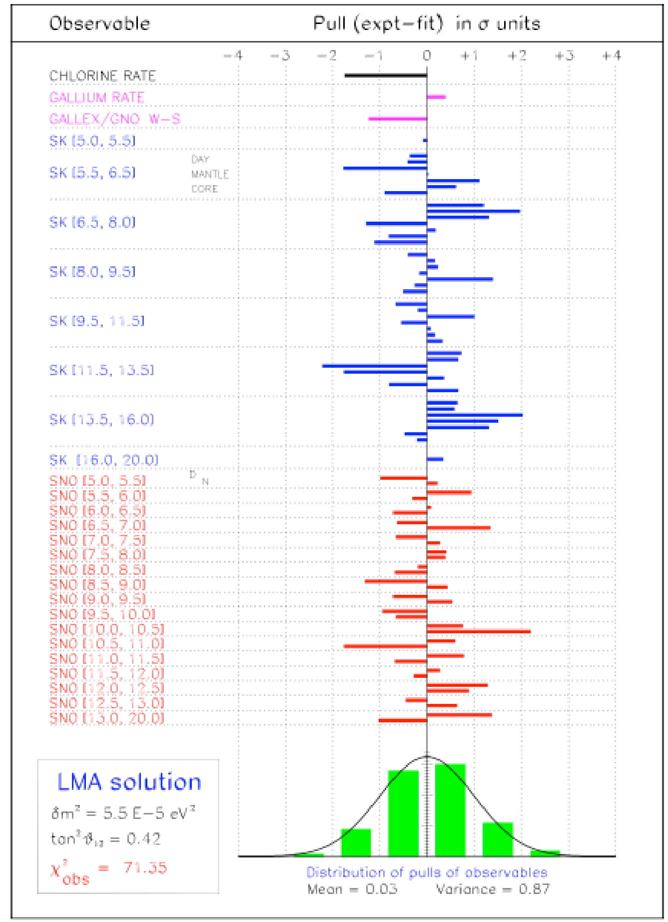
- **In formulating χ^2, linearize effects of the systematic uncertainties on data and theory comparison**

$$\chi^2_{pull} \equiv \min_{\{\xi\}} \left[\sum_{n=1}^{N} \left(\frac{R_n^{\mathrm{expt}} - R_n^{theor} - \sum (c_n^k \xi_k)}{u_n} \right)^2 + \sum_{k=1}^{K} \xi_k^2 \right]$$

  - Uncertainties u_n for each observable
- **Introduce "random" pull ξ_k for each systematic**
  - Coefficients c_k^n to parameterize effect on *nth* observable
  - Minimize χ^2 with respect to ξ_k
  - Look at contours of equal $\Delta\chi^2$

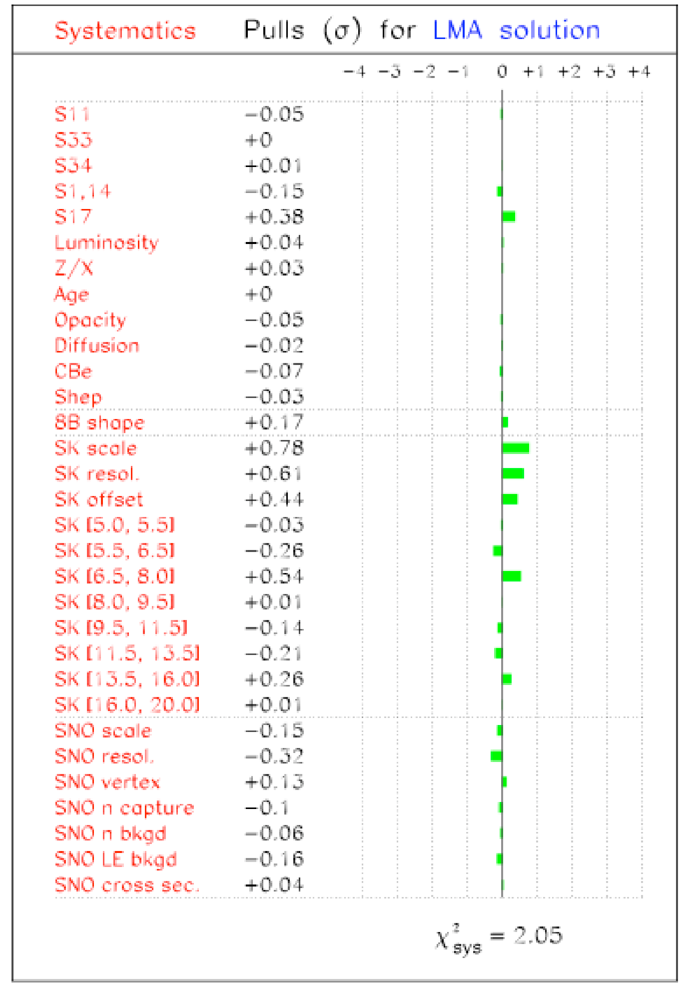
Solar Neutrino Results

- Can look at "pulls" at χ^2 minimum
  - Have reasonable distribution
  - Demonstrates consistency of model with the various measurements
  - Can also separate
    - Agreement with experiments
    - Agreement with systematic uncertainties



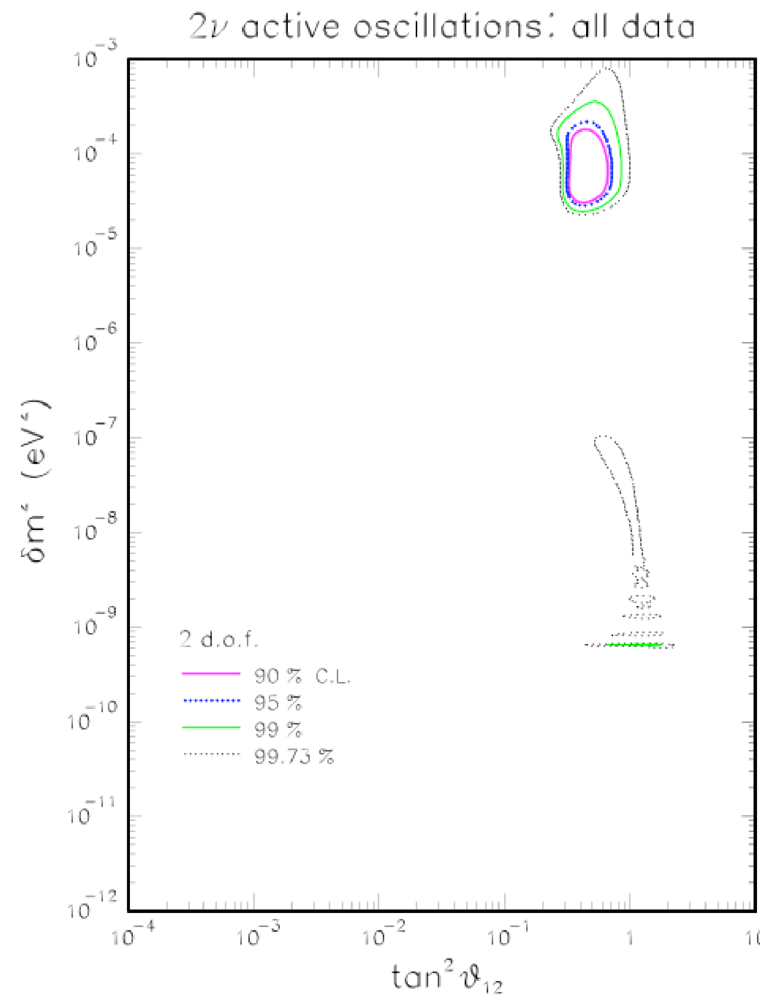
Pull Distributions for Systematics

- **Pull distributions for □$_k$ also informative**
  - **Unreasonably small variations**
  - **Estimates are globally too conservative?**
  - **Choice of central values affected by data**
    - Note this is NOT a blind analysis
- **But it gives us some confidence that intervals are realistic**



Typical Solar Neutrino Contours

- **Can look at probability contours**
  - **Assume standard χ^2 form**
  - **Probably very small probability contours have relatively large uncertainties**



Hybrid Techniques

- A popular technique (Cousins-Highland) does an "averaging" of the pdf
  - Assume a pdf for nuisance parameter g($\square$)
  - "Average" the pdf for data *x*

$$p_{CH}(x \mid \square) \equiv \square\, p(x \mid \square, \square)\, g(\square)\; d\square$$

  - Argue this approximates an ensemble where
    - Each measurement uses an apparatus that differs in parameter $\square$
      - The pdf g($\square$) describes the frequency distribution
    - Resulting distribution for *x* reflects variations in $\square$
- Intuitively appealing
  - But fundamentally a Bayesian approach
  - Coverage is not well-defined

See, for example, J. Conrad et al.

Summary

- **HEP & Astrophysics becoming increasingly "systematic" about systematics**
  - **Recommend classification to facilitate understanding**
    - Creates more consistent framework for definitions
    - Better indicates where to improve experiments
  - **Avoid some of the common analysis mistakes**
    - Make consistent estimation of uncertainties
    - Don't confuse cross-checks with systematic uncertainties
- **Systematics naturally treated in Bayesian framework**
  - **Choice of priors still somewhat challenging**
- **Frequentist treatments are less well-understood**
  - **Challenge to avoid loss of information**
  - **Approximate methods exist, but probably leave the "true frequentist" unsatisfied**