Statistical Analysis of Systematic Errors and Small Signals

Reinhard Schwienhorst
University of Minnesota

10/26/99

Outline

- Introduction
- Statistics
  - Introduction
  - Philosophies
- Confidence limit and systematic uncertainty
  - Typical Problem
  - What is a 90% confidence limit?
  - Feldman-Cousins method
  - Bayesian method
- Systematic uncertainty
  - Bayesian
  - Frequentist
  - Cousins-Highland
- Conclusions

Introduction

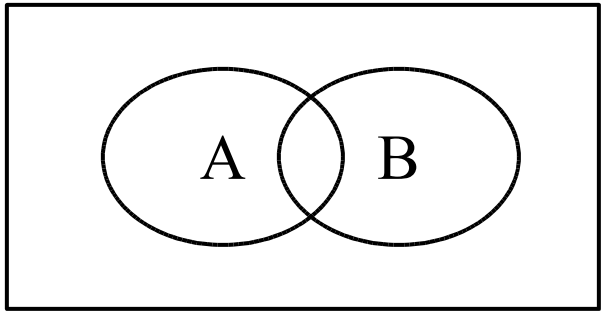
- Physics experiments are usually out to
  - Discover something
    - Find *events* that cannot be explained by the standard model
    - Find a few events above a background

      Statistics of small numbers
  - Measure something very precisely
    - Analyze many events in detail
    - Have very good control over the experiment

      Systematic uncertainty

Statistics Introduction

- Probability to observe x in a single measurement is $P(x|\mu_{true})$
  - μ_{true} stands for all parameters that describe the probability distribution function (*Poisson: mean, Gaussian: mean and sigma*)
- Bayes theorem:
  For two sets A and B:
  $P(A|B)\,P(B) = P(B|A)\,P(A)$



  - P(A) is the probability to find A (out of all the possible events in the world)
  - P(B|A) is the probability to find B in the set A

Statistics Philosophies

Probability is:

Frequentist

- The limiting relative frequency of a certain outcome:

$$P(A) = \lim_{n\to\infty} \frac{\#\text{ of outcome A in } n \text{ measurements}}{n}$$

- True values can never be determined precisely
- Includes several assumptions
  - Experiment is repeatable, paremeters don't change, each measurement has the same probability, ...

Bayesian

- Subjective:

$$P(A) = \text{degree of belief that hypothesis A is true}$$

- Intuitive definition
- Included frequentist statement
- Make statements about the probability of a true value with Bayes theorem:

$$P(\mu_{true} \mid x) \propto P(x \mid \mu_{true}) P(\mu_{true})$$

Typical Problem

- Search for events generated in a non-Standard Model process
- The number of observed events is given by

$$\boxed{n_{found} = n_{se} \times S + n_{bg}}$$

where:

  - n_{found}: the found number of events, probabilistic variable
  - n_{se}: the number of signal events, fixed but unknown
  - S: the sensitivity, with systematic uncertainty
  - n_{bg}: the number of background events due to ordinary SM processes, fixed and known
- The experiment tries to determine if n_{se} 0
- Usually a 90% confidence interval is given

What is a 90% Confidence Interval?

Frequentist

- *If I repeat an experiment many times (and create a confidence interval in each experiment), the true value μ_t will lie inside the interval 90% of the time.*
- Statement about many (hypothetical) experiments
- We (Physicists) like to argue in Frequentist terms
  - We try to convince others in Frequentist language

Bayesian

- *If I observe x_0 in a single experiment, 90% of the possible values for the true value μ_t lie inside the Bayesian interval.*
- Statement about the true value
- We (Physicists) like to think and feel in Bayesian terms
  - We form our own opinion with Bayesian intuition

Feldman-Cousins Method

- Frequentist philosophy

$$n_{found} = n_{se} \times S + n_{bg}$$

- For each possible theoretical value for n_{se}, find the distribution of n_{found} , n_{bg} and S are fixed
- For each of those distributions, find the 90% acceptance interval
  - Add n_{found} values to the interval I until $P(I|n_{se})$ 90%
  - Use the likelihood ratio $P(n_{found}|n_{se})/P(n_{found}|n_{se,best})$ to decide which values to add next
- Make a plot of n_{se} vs I (confidence belt)
- Open the box, take out the experimental result (n_{obs}) and read the confidence interval off the belt

Feldman-Cousins Method

Poisson distribution and likelihood ratio

Confidence belt

Title:
C:distribution1.eps
Creator:
ROOT Version 2.23/04
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

Title:
c1.eps
Creator:
ROOT Version 2.22/10
Preview:
This EPS picture was not saved
with a preview included in it.
Comment:
This EPS picture will print to a
PostScript printer, but not to
other types of printers.

* : Poisson distribution, $\mu=3$
: distribution of the "best" value
: ratio R

Bayesian Solution

$$n_{found} = n_{se} \times S + n_{bg}$$

- $P(n_{found}|n_{se},S,n_{bg})$ is a Poisson distribution $f_P(n_{found})$
- Likelihood to find n_{se} is

$$f(n_{se}|n_{found}) = \frac{f_P(n_{found}) f_0(n_{se})}{\int f_P(n_{found}) f_0(n_{se}) dn_{se}}$$

- Depends on the prior belief in n_{se}, expressed through $f_0(n_{se})$
  - Usually $f_0(n_{se})$=constant
- For large values of n_{se}, the result is identical to the frequentist method

What is Systematic Uncertainty?

Example: sensitivity $\boxed{S = 0.8 \pm 0.1}$

Frequentist

- Not an error
- Not clear, possibilities:

a S changes from event to event, due to variation in
  - pulseheight, position, temperature, …

b S changes from experiment to experiment
  - S determined by MC
  - Uncertainty due to finite # of MC events

c Uncertainty about S (?)

Bayesian

- Not an error
- Degree of belief in S
- *I think S is 0.8, but I am not sure*
- It expresses how well we understand our detector

Bayesian Treatment of Systematic Uncertainty

$$n_{found} = n_{se} \times S + n_{bg}$$

- Solve as before; also integrate over the uncertainty in S:

$$f(n_{se} \mid n_{found}) = \frac{\int f_P(n_{found} \mid n_{se}, S) f_0(n_{se}, S) dS}{\iint f_P(n_{found} \mid n_{se}, S) f_0(n_{se}, S) dn_{se} dS}$$

- $f_0(n_{se},S)$ includes a distribution for S
  - in our example a Gaussian with mean 0.8 and width 0.1
- S is called an influence variable
  - it affects the result, but we don't care what it is

Problems with the Bayesian Treatment

- What should the prior $f(n_e,\mu_S,\sigma_S)$ be?
  - The experimenter should decide
- What is $f(S_{mean},\sigma_S|\mu_S,true,\sigma_{S,true})$?
  - Typically, one sets $\sigma_S = \sigma_{S,true}$ and assumes a Gaussian distribution for $f(S_{mean}|\mu_S,true,\sigma_{S,true})$
- What is the interpretation of $f(S_{mean},\sigma_S|\mu_S,true,\sigma_{S,true})$?
  - Degree of belief
  - Level of trust in S_{mean}

Frequentist Treatment of Systematic Uncertainty

- Should we even do this?
- If we don't know the sensitivity of our detector, we should
  - improve our calibration
  - make more tests
  - generate and analyze more MC events
  - not publish our results (?)
- If we know the sensitivity of our detector, it is not a probabilistic variable and we should use the (known) value.

Treatment of Frequentist Systematic Uncertainty a

a Event-to-event variation interpretation:

- Use calibration data to get the response variations of detector elements
  - Testbeam
  - Cosmic ray tests
  - Subsets of real data
- Put variations in pulseheight, position resolution , … into the MC simulation
- Use MC simulation to get the average S
- Generate the confidence belt according to

$$n_{obs} = n_e \times S + n_{bg}$$

Treatment of Frequentist Systematic Uncertainty b

- Global variation interpretation
  - Uncertainty in flux normalization, # of MC events
- The probability distribution for S is a Gaussian with mean μ and width σ
- In the experiment, we obtain one measurement for S and an uncertainty

  Estimate mean and width from the experiment

  Create a multidimensional confidence belt to get confidence intervals for n_{se} and S (covariance matrix)

– But we don't care about a limit for S
  - quote the worst-case interval for n_{se}
  - or average (integrate) over S
  - or ?

Another View of Systematic Uncertainty

- Conditions change through the run of an experiment
  - The experiment is practically not "repeatable"
- Event-to-event uncertainties are usually built into the MC
  - GEANT, detector simulation
- Control data is used to estimate the global uncertainty in S
  - In E872, use ν_μCC events to check the ν_τCC signal
  - If control data agrees with control MC (based on statistical tests), the systematic uncertainty is at most small
    - uncertainty due to # of control events is not a systematic error
    - generate enough control data to make the uncertainty negligible
  - But:
    - comparing MC and data is really a matter of belief
    - how do we deal with flux normalization etc.?

Cousins-Highland solution

- Frequentist treatment of the Poisson variable
- Bayesian treatment of the systematic uncertainty
- Average (integrate)

$$\boxed{n_{obs} = \int \left(n_e \times S + n_{bg} \right) f(S) dS}$$

- Then use Frequentist methods to get a confidence interval

Coverage

- Frequentist concept
- For a true value inside my 90% confidence interval:
  - Generate many experiments according to the probability distribution.
  - For each experiment, find the confidence interval.
  - The sum of all probabilities for intervals that contain the true value should be at least 0.9 (the intervals should cover the true value)
- Simplified: *The true value will be covered by the confidence interval in 90% of the experiments trying to measure it.*
- Bayesians don't care about coverage

Conclusions

- Systematic errors should be minimized as much as possible
- If they are non-negligible, they should be handled (interpreted) with caution
- Different methods exist for dealing with small signals
- The statistical analysis of small signals is under debate (Frequentist Bayesian)
- There will be a workshop at CERN in January (http://www.cern.ch/CERN/Divisions/EP/Events/CLW/)