Systematic Uncertainty

Physics 252C - Lecture 11
Prof. John Conway
what is a systematic uncertainty?

- if the estimator of the physical quantity you are trying to measure depends on other quantities that you do not know or know only imperfectly, you have **systematic errors**

- in general, we can say that systematic errors arise due to imperfect **modeling** of our observables

- systematic errors are, too often, an afterthought in many analyses

- you should develop the habit of thinking about what systematic errors you might encounter from the **very beginning of your analysis** and how to deal with them

- the first step is to measure them, then incorporate them
systematic effects

- once you have identified a potential source of systematic error, you need to ask the question: “how does a change due to this effect alter the result I get for my estimator?”

- in short we need to evaluate \( \frac{\Delta \hat{\alpha}}{\Delta x} \) where the variable \( x \) represents the effect

- then (and only then!) we can evaluate the systematic error on \( \alpha \) due to the uncertainty in \( x \):

\[
\sigma_\alpha = \left( \frac{\Delta \hat{\alpha}}{\Delta x} \right) \sigma_x
\]

- sometimes the hardest part is figuring out \( \sigma_x \)
evaluating the derivative

- there is a trap here: how do we really evaluate the derivative?
- it's different for different effects
- sometimes we simply know (because someone tells us, perhaps?) that the uncertainty in our parameter due to some effect is a certain relative error
  - example: integrated luminosity uncertainty, with respect to a cross section
- in many cases we can represent the effect as a parameter in our likelihood... then all we need to know is the uncertainty in the parameter
“nuisance” parameters

- clearly we don’t care about many of our parameters; these are “nuisance” parameters
- in a Bayesian treatment, we need to remove them in some way to obtain a posterior pdf in the parameter of interest
- can represent systematics as likelihood parameters
- but....then what?
- answer next Monday, and beyond...
- for now let us assume we are simply interested in determining the change in nuisance parameters due to systematic effects
nuisance parameters

- this is the way to deal with systematics, whenever possible (trust me)
- people are a bit confused at times about which are the nuisance parameters, and which are something else
- there is no something else!
- any parameter not of interest can be regarded as a nuisance parameter, and treated as a systematic error
- this is another reason to focus on the likelihood as the core of an analysis
- we shall henceforth refer to the parameter(s) of interest as $\alpha$, and the nuisance parameters as $\beta$
measuring the derivatives

- what we care about now is \( \frac{\Delta \beta}{\Delta x} \)
- make a change in \( x \) and then see how beta changes
- have to be careful
- need to make either very accurate or multiple measurements:
- perversely shown for negative slope...does that matter?
pseudoexperiments for systematics

- to generate a distribution like the one on the previous slide, what do we do?
- make a change $\Delta x$, then generate a fake (pseudo)data sample
- then determine the value of $\beta$ that results, and record it
- repeat many times
- must make the plot to get an idea of accuracy!
systematic effects to consider

- integrated luminosity
- trigger efficiency
- identification efficiency
- geometric acceptance
- energy/momentum scale
- energy/momentum resolution
- energy/momentum nonlinearity
- background rate
- parton density functions
- choice of $Q^2$ scale
- ISR/FSR
- Monte Carlo statistics
- theoretical uncertainty
example analysis: $H \rightarrow \tau\tau$ in CDF

- in MSSM can get enhanced production of higgs
- select events with $e+\tau$, $\mu+\tau$, $e+\mu$
- the “money” plot: visible mass distribution of tau pairs
- systematic errors include all the ones I listed on the previous slide!
- systematics can take ~75% of your time to do an analysis
• all of these are enhanced like tan^2\beta
• can look for \( \phi \rightarrow bb \) or \( \tau\tau \)
• Tevatron and LHC have sensitivity
**integrated luminosity uncertainty**

- we measure integrated luminosity using a very forward calorimeter to essentially measure the total $p\bar{p}$ or $pp$ inelastic cross section

- most collision processes give very forward particles and jets

- if we know the total cross section, from the observed rate can infer the (integrated) luminosity

- this sets the scale for what is $1 \text{ fb}^{-1}$

- CDF: Cerenkov Luminosity Calorimeter; 6% unc.

- CMS: Totem will measure elastic and inelastic cross sections, which are related by the optical theorem
integrated luminosity uncertainty
trigger efficiency

- need to know what the efficiency is: measure it!
- residual uncertainty in measurement leads to a systematic uncertainty which adds in quadrature with the integrated luminosity uncertainty
- difficult to simulate trigger performance
- use prescaled “backup” triggers with reduced requirements
- study each trigger requirement individually
- best strategy overall: cut well within the trigger turn-on thresholds
- Level 1, Level 2, Level 3 ...
trigger efficiency

- example: track trigger efficiency in CDF
- plot as efficiency versus $1/p_T$

![Graph showing efficiency versus $1/p_T$]

- $\chi^2 / \text{ndf} = 70.18 / 39$
- Prob = 0.001607
- p0 = 0.9874 ± 0.0012
- p1 = -0.1139 ± 0.0210
- p2 = 0.2049 ± 0.0006
- p3 = -0.006963 ± 0.000371
identification efficiency

- require electrons, muons, taus in the example analysis; simulation is not perfect

- electrons and muons: can use $Z \rightarrow ll$ events, place tight cuts on one leg, then use mass cut
  - other leg is "tagged" and can go in denominator
  - place cuts on tagged lepton to get numerator

- taus: more difficult, as neutrinos make it hard to tag taus

- use large $W \rightarrow \tau \nu$ sample to cross check efficiency

goal: lepton ID scale factors (data/MC) with uncertainties
energy scales

• momentum scale (tracking)
  ▶ can use $J/\psi$, $\Upsilon$, $Z$ masses to calibrate

• energy scale (em calorimetry)
  ▶ use tagged electrons from $Z$ masses to calibrate

• energy scale (had calorimetry)
  ▶ use $\Upsilon +$ jet events, require $p_T$ balancing

• for taus have the additional problem that $\pi^0$ reconstruction not perfectly modeled

for the $H \rightarrow \tau\tau$ analysis this is the largest uncertainty!
energy nonlinearity

- calorimetry may have non-linear response
- must measure $E_{\text{cal}}$ versus $E_{\text{true}}$ and derive correction function
- residual uncertainty on $E$ leads to systematic uncertainties on
  - acceptance (due to threshold cuts)
  - kinematic distribution shape
background rates

- can share systematic uncertainties with signal
- sometimes we predict background using some sort of extrapolation from background-rich/signal poor to background-poor/signal rich
- residual uncertainty on background is a systematic error
- background rate itself is a nuisance parameter
- for H\rightarrow\tau\tau analysis we fit for the background rate...
Monte Carlo statistics

- In a binned likelihood fit, we endeavor to add together signal and background to get observed
- The data statistics and the model statistics, in effect, add in quadrature
- Endeavor to generate so much MC that the increased error is negligible...
- 10x more MC than observed means 5% additional error
- Multiple bins reduces this, right?
theoretical uncertainties

- our MC programs imperfectly simulate nature
- analytic cross section calculations may be used for the “hard” process...or not
- actual production rate depends on parton densities (CTEQ, MRST, ...) which provide uncertainties
- PDF uncertainties typically affect cross section, mainly, but can in principle affect shape
- gluon densities have large uncertainties; b parton pdf is very uncertain
- $Q^2$ scale: evaluate acceptance/shape of kinematic distributions, altering scale; similarly with ISR/FSR