Wish List of ATLAS & CMS

Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people:
Alex Read, Bill Quayle, Luc Demortier, Robert Thorne
Glen Cowan ,Kyle Cranmer
& Bob Cousins

Thanks also to Ofer Vitels

Fred James
In the first Workshops on Confidence Limits
CERN & Fermilab, 2000

- Many physicists will argue that Bayesian methods with informative physical priors are very useful

WHAT I WOULD LIKE TO SEE:

1. PHYSICISTS LEARN THE VOCABULARY OF STATISTICS
2. ASSUMPTIONS, METHODS, APPROXIMATIONS CLEARLY SPECIFIED IN PUBLICATIONS
3. FELDMAN/COUSINS IN ALL SEARCHES
4. BAYESIAN DECISION THEORY IN POLICY DECISIONS

The Target Audience for this Talk

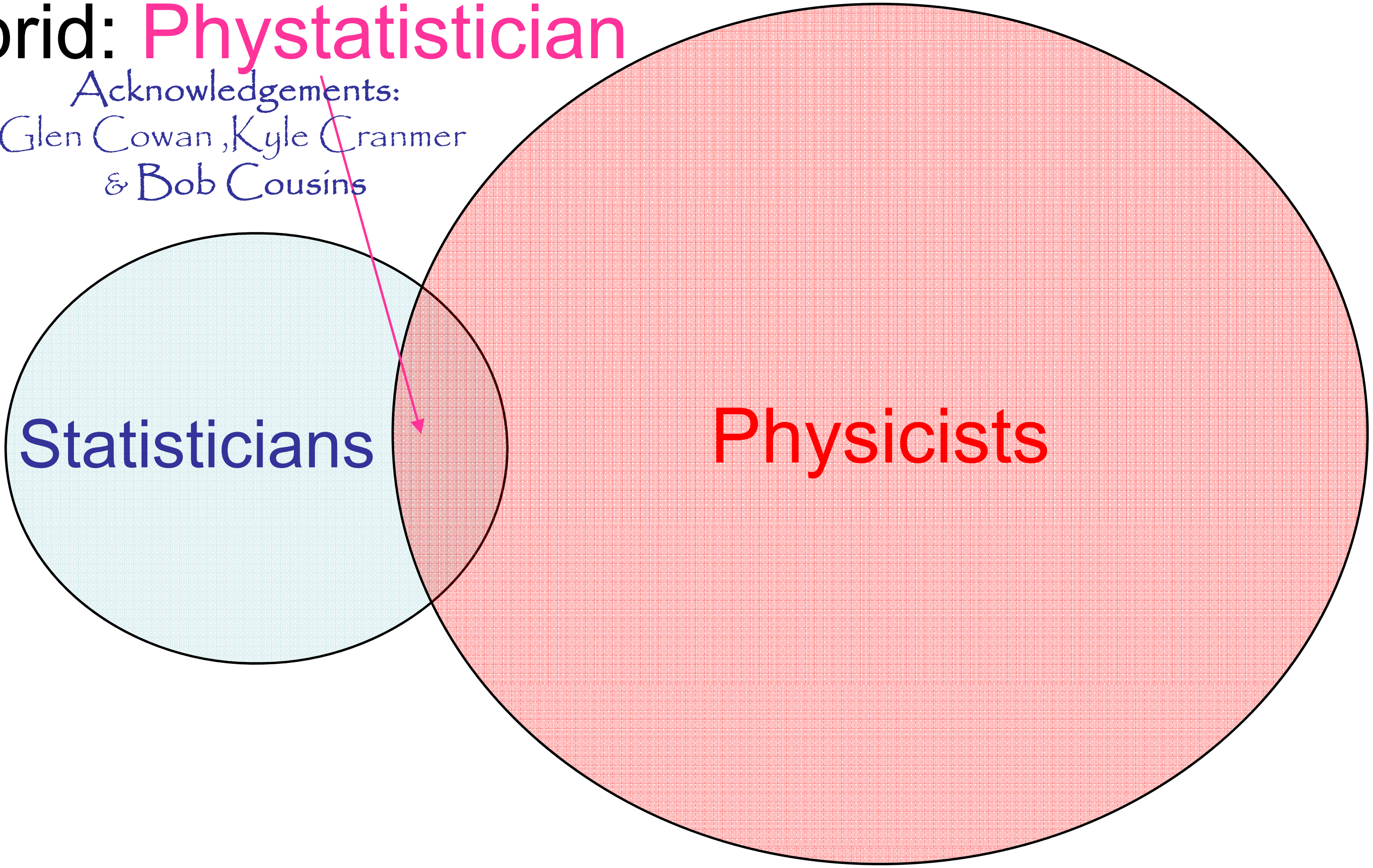
- A typical concern e-mail:
  - *"we have been recently discussing how to incorporate systematic uncertainties in our significances & thereby discovery/exclusion contours. Other issues involve statistics to be used or statistical errors, in particular keeping in mind that each group*
    should do this investigation in a way such that we don't have problems later to combine their results"

Wish List – Submitted to Whom?

$$Physicists \cup Statisticians \approx 200$$

A hybrid: Phystatistician

Acknowledgements:
Glen Cowan ,Kyle Cranmer
& Bob Cousins



The Phystatisticians Club

- A pre-requisite: Reading of original professional statisticians papers and books
- First in the list: The original Kendall and Stuart (age 47 now)

M. Kendall & A. Stuart, **"The Advanced Theory of Statistics" vol 2, ch 24**

- That is certainly not last in the list, there is a lot of modern stuff…

Now consider the likelihood ratio

$$l = \frac{L(x \mid \theta_{r0}, \hat{\hat{\theta}}_s)}{L(x \mid \hat{\theta}_r, \hat{\theta}_s)}. \quad (24.4)$$

. . .

Intuitively, l is a reasonable test statistic for H_0: it is the maximum likelihood under H_0 as a fraction of its largest possible value, and large values of l signify that H_0 is reasonably acceptable. The critical region for the test statistic is therefore

$$l \leqslant c_\alpha, \quad (24.6)$$

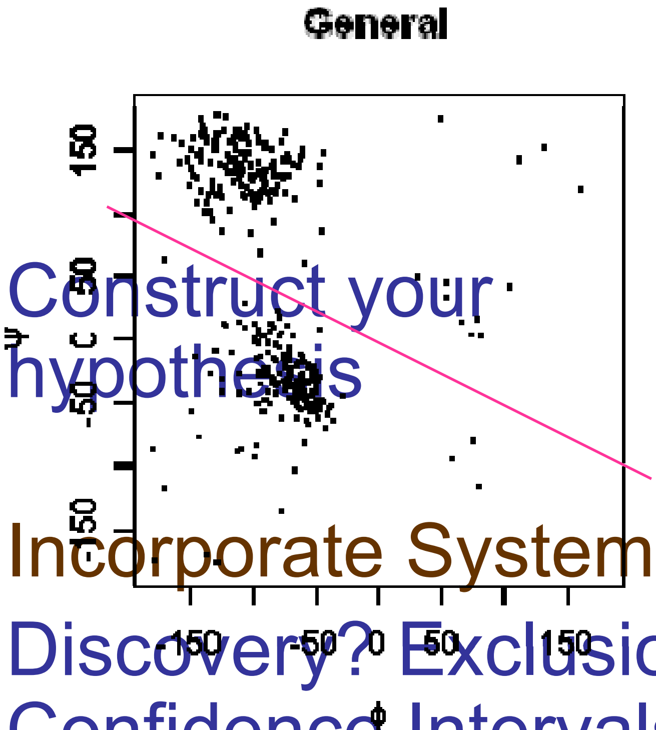
where c_α is determined from the distribution $g(l)$ of l to give a size-α test, i.e.

$$\int_0^{c_\alpha} g(l)\,dl = \alpha. \quad (24.7)$$

The Stages of a Physics Analysis

- Modeling of the underlying processes
- Selection of preferred data
- Fitting & Testing
- Uncertainties (Statistic, Systematic)
- Interpreting the results

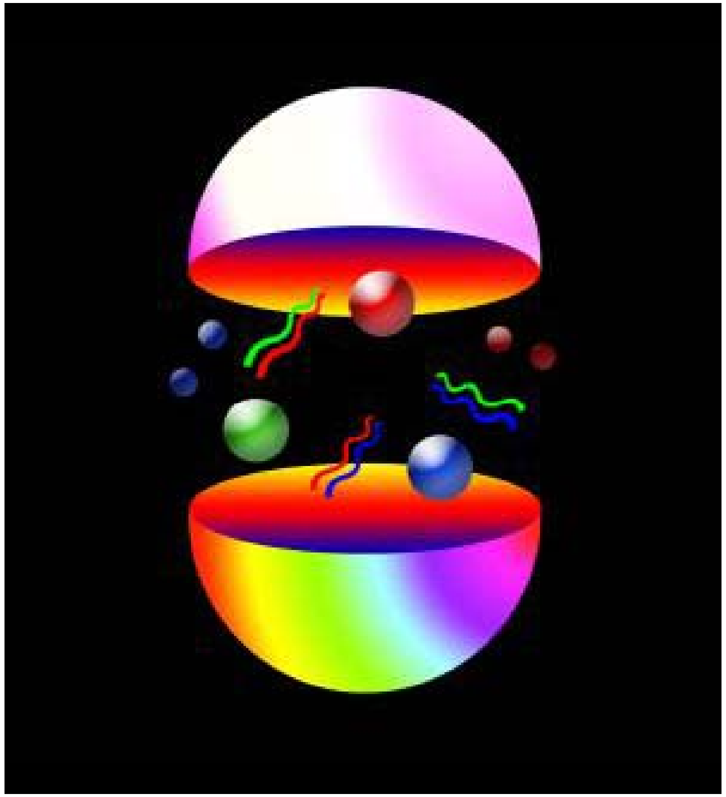
- MC… Pythia, Herwig



- Construct your hypothesis
- Incorporate Systematics
- Discovery? Exclusion? Confidence Intervals….

Modeling of the Underlying Processes

- The probability density to find in the proton a parton i, which carries a fraction x of the Proton momentum is dependent on the energy scale Q^2 and is given by $f_i(x;Q^2,\alpha_s(Q^2))$



A Possible "Global Analysis" Paradigm

John Pumplin

1. Parameterize the *x*-dependence of each flavor (u,d,s,g…) at some Q_0^2 (~ 1 GeV) with
2. Compute PDFs $f_i(x;Q^2,\alpha_s(Q^2))$ at all $Q^2 > Q_0^2$ by running
3. Compute cross sections for Deep Inelastic

Scattering, Drell-Yan, Inclusive Jets,. . . using

QCD perturbation theory

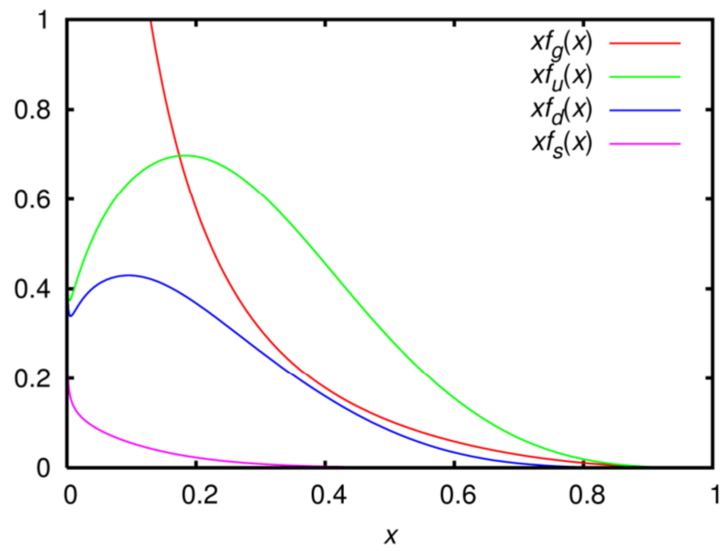
4. Compute "χ^2" measure of agreement between

predictions and measurements:

$$\chi^2 = \sum_i \left(\frac{\text{data}_i - \text{theory}_i}{\text{error}_i} \right)^2$$

5. Vary the parameters in $f_i(x;Q^2,\alpha_s(Q^2))$ to minimize χ^2,

yielding Best Fit pdfs: CTEQ6.1, MRST,. . .



$$\sigma(pp \to X) \sim x_i f_{i(x_i,Q^2,\alpha_s(Q^2))} \otimes \sigma(ij \to X)_{(x_i,x_j,\alpha_s(Q^2))} \otimes x_j f_{j(x_j,Q^2,\alpha_s(Q^2))}$$

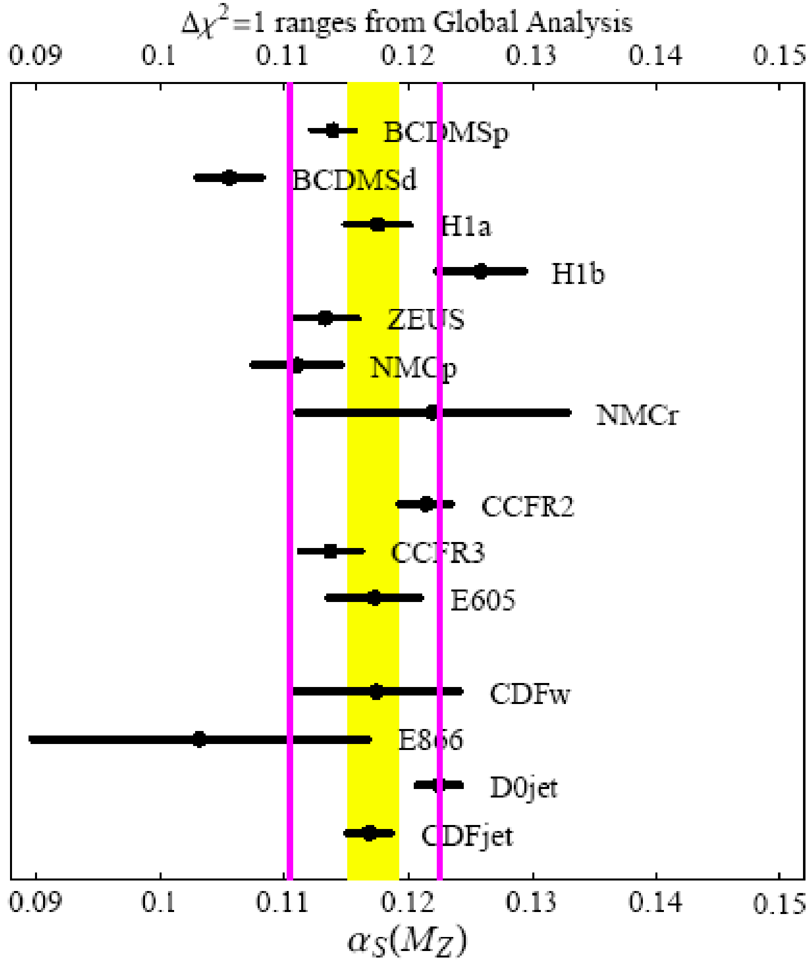
The Global Fit Results

- χ^2/dof=2659/2472 with 30 free parameters overall (MRST).
  - p-value <1%
- For most of the individual experiments fit the χ^2/dof~1 or a bit less.
- Normally one would take $\Delta\chi^2$=1 to estimate the error

Statistical

Either Inflate $\Delta\chi^2$ or Equivalently Inflate the Errors

- The size of the errors and scatter of the points is unrealistic
- If "what you put" in is "what you get" out – there must be a problem with "what you put" in.



$\Delta\chi^2$~100

CTEQ6 $(\Delta\chi^2 = 100)$

$$\alpha_S(M_Z^2) = 0.1165 \pm 0.0065$$

A Statistical Analysis

- What's happening?

- The $\Delta\chi^2=1$ rule assumes that statistical and systematic errors are understood and known.

- Stump (Phystat 2003) argues: What we have are estimates on the uncertainties, not the true ones… The increase of χ^2 if the estimators are biased or wrong might be bigger than 1!
- He concludes: "We find that alternate pdfs that would be…unacceptable differ in χ^2 by an amount of order……… 100!!! "

- Unacceptable is a very vague statement

A Wish

- It would be nice to have a more systematic way of accounting for this, e.g. a modified definition of goodness of fit to account for non-Gaussian nature of errors, a quantitative way of accounting for theoretical errors, etc, but all these are very difficult, therefore allowing for a larger than usual increase in $\Delta\chi^2$ seems like a "sensible" way of accounting for the full effect. Robert Throne

Interpreting The Results
Significance or Exclusion

- So there are two alternate questions (Do not confuse between them):
  - Did I or did I not establish a discovery
    - Goodness of fit (get a p-value based on LR)
  - How well my alternate model describes this discovery
    - measurement…., here one is interested in a confidence interval $[m_l,m_u]$ (more to come)
  - In the absence of signal, derive an upper limit

Hypothesis Testing - The Ingredients

- LHC Physics Community = LEP+CDF+D0+BaBaR+Belle…
- That is (CLs+Hybrid)+Bayesian+Frequentist (a' la F&C)
- So the only way out is to do it all….
  But, in a way
  One leads to another….. (conceptually)

Testing of Hypotheses

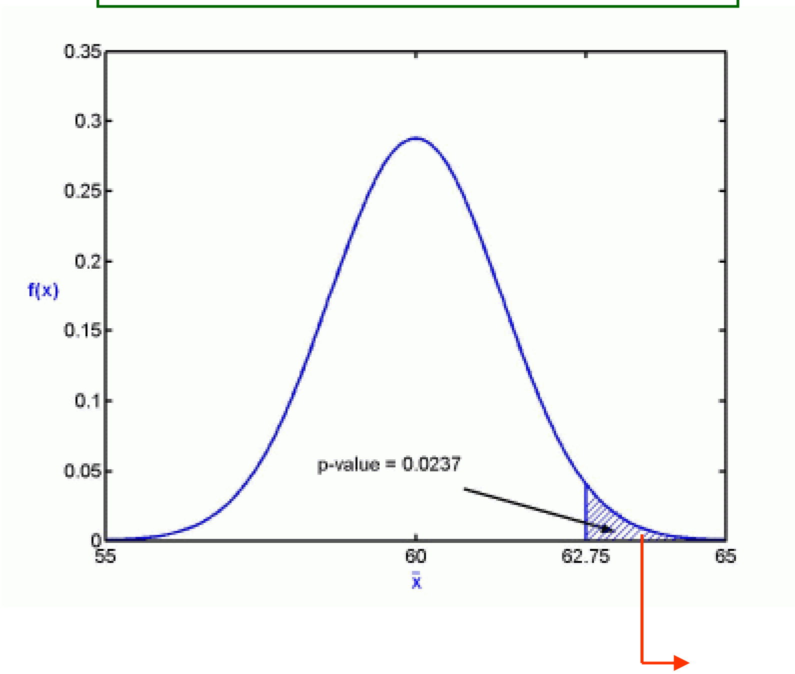
- Given an hypothesis H_0 (Background only) one wants to test against an alternate hypothesis H_1 (Higgs with mass m)
- One (very good) way is to construct a test statistics Q

$$Q(m) = \frac{L(H_1)}{L(H_0)} = \frac{L\left(s(m)+b\right)}{L(b)}$$

Basic Definitions: p-Value

- A lot of it is about a language…. A jargon
- Discovery…. A deviation from the SM - from the background only hypothesis…
- p-value = probability that result is as or less compatible with the background only hypothesis
- Control region α (or size α) defines the **significance**
- If result falls within the control region, i.e. **p< α** BG only hypothesis is rejected
  →A discovery

- The pdf of Q….

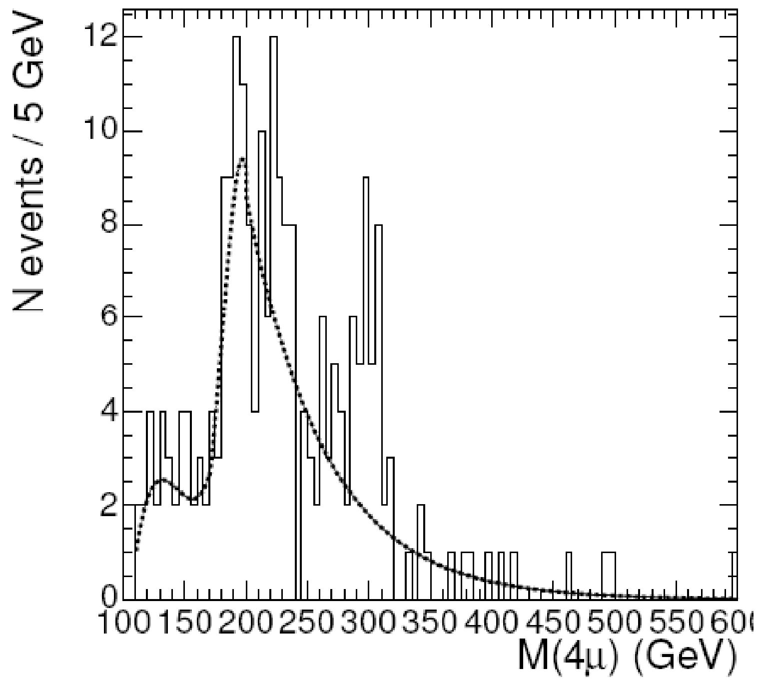


Control region
Of size α

Look Elsewhere Effect

- One way to reduce the effect is to use a prior on the Higgs mass (based for example on EW measurements)… but do we want to use priors on our parameters of interest… ?
- **Take more data**
  - Hopefully the LHC situation is better than LEP in that sense
- Do not rule out the following:
  - A quick way to trace a look elsewhere effect before asking the statisticians is to literally LOOK ELSEWHERE, i.e. consult between ATLAS and CMS…..
  - Hiding does not always do you good…

See talk of Alexey Drozdetskiy



CMS TDR App. A

Why 5σ?

- A simple calculation reveals a possible explanation:
- Having 100 searchesx100 resolution bins→ False discovery rate of 5σ is $(10^4)x(2.7x10^{-7})=0.27\%$ ➔ That's effectively a 3σ
  (see also Feldman, Phystat05)

See Kyle's Talk

- The related question would be:
  - How many σ are needed to claim a discovery?….
  - Is there a problem in seeking for an effect at a tail of the pdf (if not a signal, it might be a tail fluctuation…)?
  - Or is it not the right question…. Is our procedure for discovery (p-values) correct considering the involved playground (SM vs SM+Higgs vs SUSY….)

Sequential Analysis – Stopping Rules

- Looking 1 time at data (could be blind till then) and trying to determine best time at which we should look at them => fixed sample size analysis
- Looking at data as they come, allow early stopping, but try to adapt statistical methods => sequential analysis
- I was almost convinced by Renaud Bruneliere that by predefining a stopping rule a' la Wald (1945) , we would achieve a discovery with half the luminosity
- **Should we consider adopting a stopping rule for Higgs discovery, or am I completely out of my mind?**
- ~ " I think, I shall wait till I am retired to try and understand stopping rules" (Bob)
- But a better wish would be (G. Cowan):
  Can we have a rough guide that will tell us by how much our p-value is increased as a result of the fact that we have already looked at the data a few times before and got no satisfactory significance?
  (something like a look elsewhere effect in the time domain….)

Basic Definitions: Confidence Interval & Coverage

- Say you have measurement m_{meas} of m with m_t being the unknown true value of m
- Assume you know the pdf of $p(m_{meas}|m)$
- Given the measurement you deduce somehow that there is a 90% Confidence interval $[m_1,m_2]$....
- The misconception; Given the data, the probability that there is a Higgs with a mass in the interval $[m_1,m_2]$ is 90%.
- The correct statement: In an ensemble of experiments 90% of the obtained confidence intervals will contain the true value of m.

Basic Definitions

- **Confidence Level:** A CL of 90% means that **in an ensemble of experiments, each producing a confidence interval, 90% of the confidence intervals will contain s**
- Normally, we make **one** experiment and try to **estimate from this one experiment** the **confidence interval** at a specified CL% Confidence Level….
- If in an ensemble of (MC) experiments the true value of s is covered within the **estimated** confidence interval , we claim a coverage
- If in an ensemble of (MC) experiments our **estimated** Confidence Interval fail to contain the true value of s, 90% of the cases (for every possible s) we claim that our method undercovers

A Word About Coverage

- The basic question will come again and again: WHAT IS THE IMPORTANCE OF COVERAGE for a Physicist?

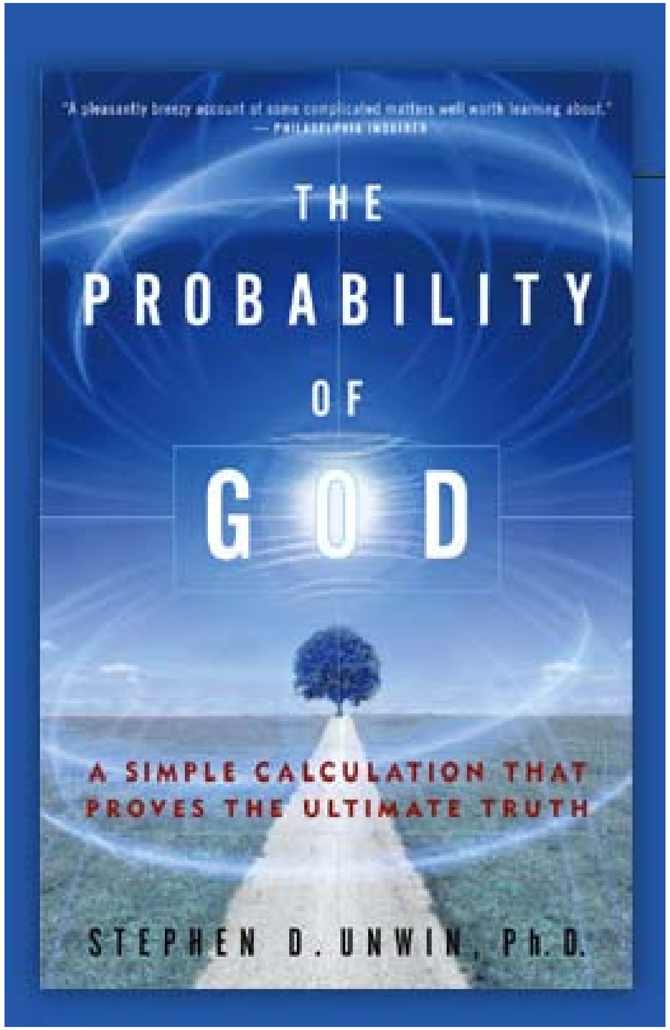
- The “ problem” : Maybe coverage answers the wrong question…
you want to know what is the probability that the Higgs boson exists and is in that mass range… So you can either educate the physicists that your exclusion does not mean that the probability of the Higgs Boson to be in that mass range is <5%... or you try to answer the right question!

What is the Right Answer?

- The Question is:
- Is there a Higgs Boson?
- Is there a God?

$$P(God \mid Earth) = \frac{P(Earth \mid God)P(God)}{P(Earth)}$$

- In the book the author uses
  - "divine factors" to estimate the P(Earth|God),
  - a prior for God of 50%
- He "calculates" a 67% probability for God's existence given earth…
- In Scientific American July 2004, playing a bit with the "divine factors" the probability drops to 2%...



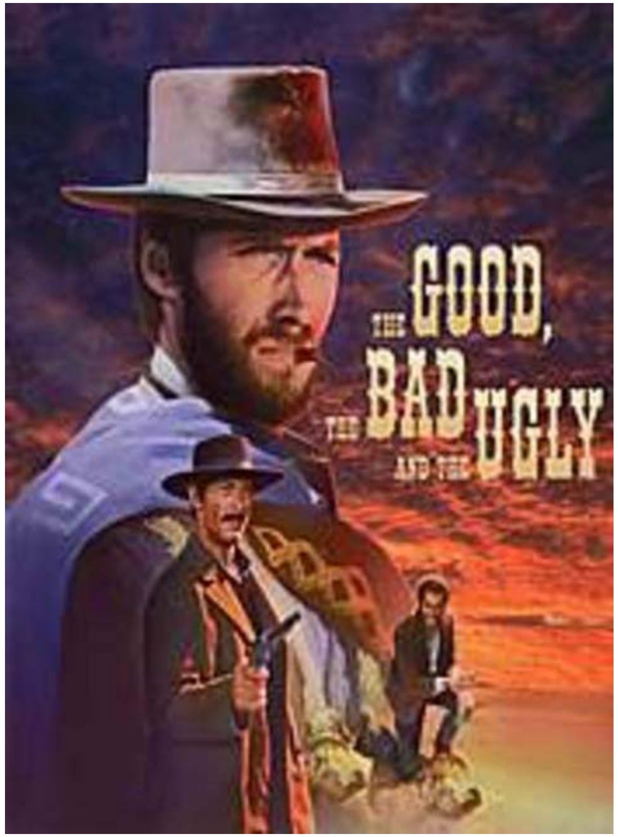
What is the Right Question

- Is there a Higgs Boson? What do you mean?
  Given the data , is there a Higgs Boson?
- Can you really answer that without any a priori knowledge of the Higgs Boson?
  Change your question: What is your degree of belief in the Higgs Boson given the data… Need a prior degree of belief regarding the Higgs Boson itself…

$$P(Higgs \mid Data) = \frac{P(Datas \mid Higgs)P(Higgs)}{P(Data)} = \frac{L(Data \mid Higgs)\pi(Higgs)}{\int L(Data \mid Higgs)\pi(Higgs)d(Higgs)}$$

- If not, make sure that when you quote your answer you also quote your prior assumption!
- The most refined question is:
  - Assuming there is a Higgs Boson with some mass m_H, how well the data agrees with that? $P(Data \mid Higgs)$
  - But even then the answer relies on the way you measured the data (i.e. measurement uncertainties), and that might include some pre-assumptions, priors!

Systematics



Why “download” only music?
“Download” original ideas as well…. ☺

Nuisance Parameters (Systematics)

- There are two related issues:
  - Classifying and estimating the systematic uncertainties
  - Implementing them in the analysis
- The physicist must make the difference between cross checks and identifying the sources of the systematic uncertainty.
  - Shifting cuts around and measure the effect on the observable…
    Very often the observed variation is dominated by the statistical uncertainty in the measurement.

Reporting Errors

- A common habits is to combine all systematic errors in quadrature and then combine those in quadrature with the statistical errors and report a result

- This is a bad habit
  Results should be reported by at least separating class 1 errors from the rest… Let the reader decide….

- From ZEUS (with $\Delta\chi^2_{eff} = 50$)

$$\alpha_S(M_Z^2) = 0.1166 \pm 0.0008(\mathrm{uncor}) \pm 0.0032(\mathrm{corr}) \pm 0.0036(\mathrm{norm}) \pm 0.0018(\mathrm{model})$$

Subjective Bayesian is Good for YOU

Thomas Bayes (b 1702)
a British mathematician and Presbyterian minister

The Bayesian Way

See also Wade Fisher talk

$$p(\theta | x) = \frac{L(x|\theta)\pi(\theta)}{\int L(x|\theta)\pi(\theta)d\theta}$$

- Can the model have a probability?
- We assign a degree of belief in models parameterized by θ
- Instead of talking about confidence intervals we talk about credible intervals, where p(θ|x) is the credibility of θ given the data.

Treatment of Systematic Errors, the Bayesian Way

- Marginalization (Integrating) (The C&H Hybrid)
  - Integrate L over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others...)
  - Consistent Bayesian interpretation of uncertainty on nuisance parameters

- Note that in that sense MC "statistical" uncertainties (like background statistical uncertainty) are systematic uncertainties

Tom Junk

Integrating Out The Nuisance Parameters (Marginalization)

$$p(\theta,\lambda \mid x) = \frac{L(x \mid \theta,\lambda)\pi(\theta,\lambda)}{\int L(x \mid \theta,\lambda)\pi(\theta,\lambda)d\theta d\lambda} = \frac{L(x \mid \theta,\lambda)\pi(\theta,\lambda)}{Normalization}$$

- Our degree of belief in θ is the sum of our degree of belief in θ given λ (nuisance parameter), over "all" possible values of λ

$$p(\theta \mid x) = \int p(\theta,\lambda \mid x)d\lambda$$

Priors

$$P(\theta | data) \sim \int L(data | \theta, \lambda) \pi(\lambda) d\theta d\lambda$$

- A prior probability is interpreted as a description of what we believe about a parameter preceding the current experiment
  - **Informative Priors**: When you have some information about λ the prior might be informative (Gaussian or Truncated Gaussians…)
    - Most would say that subjective informative priors about the parameters of interest should be avoided ("….what's wrong with assuming that there is a Higgs in the mass range [115,140] with equal probability for each mass point?")
    - Subjective informative priors about the Nuisance parameters are more difficult to argue with
      - These Priors can come from our assumed model (Pythia, Herwig etc…)
      - These priors can come from subsidiary measurements of the response of the detector to the photon energy, for example.
      - Some priors come from subjective assumptions (theoretical, prejudice symmetries….) of our model

Priors – Uninformative Priors

- **Uninformative Priors**: All priors on the parameter of interest should be uninformative….
  IS THAT SO?
  Therefore flat uninformative priors are most common in HEP.
  - When taking a uniform prior for the Higgs mass [115, ∞]… is it really uninformative? do uninformative priors exist?
  - When constructing an uninformative prior you actually put some information in it…
- **But** a prior flat in the coupling *g* will not be flat in $\sigma \sim g^2$
  Depends on the metric!
  (→ try Jeffrey Priors)
- Moreover, flat priors are improper and lead to serious problems of undercoverage (when one deals with >1 channel, i.e. beyond counting, one should AVOID them

–See Joel Heinrich Phystat 2005

Choice of Priors

- A.W.F. Edwards: "Let me say at once that I can see no reason why it should always be possible to eliminate nuisance parameters. Indeed, one of the many objections to Bayesian inference is that is always permits this elimination."

  Anonymous: "Who the ---- is A.W.F. Edwards…" http://en.wikipedia.org/wiki/A._W._F._Edwards

- But can you really argue with subjective informative priors about the Nuisance parameters
  (results of analysis are aimed at the broad scientific community.. See talk by Leszek Roszkowski constrained MSSM)

- Choosing the right priors is a science by itself

- Should we publish Bayesian (or hybrid) results with various priors?
- Should we investigate the coverage of Bayesian (credible) intervals?

- Anyway, results should be given with the priors specified

Reference Analysis

- Q: What is our prior "degree of belief"?
- Q: What will our posterior "degree of belief" be if our prior knowledge had a minimal effect (relative to the data) on our final inference?

(Quoted in **L. Demortier**, Phystat 05,)

- Reference prior function: A mathematical description where data best dominate the prior knowledge so the Bayesian inference statements only depends on the assumed model and available data (**Bernardo, Berger**)

Prescription Example: Cross Section Measurement

L. Demortier, Phystat 05

$$p(n\,|\,\sigma,\epsilon,b) \;=\; \frac{(b+\epsilon\sigma)^n}{n!}\,e^{-b-\epsilon\sigma}$$

$$\pi(\sigma;\varepsilon,b)=\pi_u(\sigma\,|\,\varepsilon,b)\pi(\varepsilon,b)$$

- A simple model
  (1 par of int+2 Nuisance par)
- Informative subjective priors for the efficiency and b (multiplication of Gamma distributions)
- Start with the Jeffrey's Prior

The ref prior for 1 variable is reduced to Jefrey's prior

$$\pi_{\mathcal{J}}(\sigma\,|\,\epsilon,b) \;\propto\; \frac{\epsilon}{\sqrt{b+\epsilon\sigma}}$$

- Normalize w.r.t. a nested set for σ: [0,u] and take a limit

$$\pi_u(\sigma\,|\,\epsilon,b) \;=\; \frac{\epsilon}{\sqrt{b+\epsilon\sigma}}\;\frac{\mathbf{1}(u\geq\sigma)}{2\sqrt{b+\epsilon u}\,-\,2\sqrt{b}}$$

$$\pi(\sigma\,|\,\epsilon,b) \;=\; \lim_{u\to\infty}\frac{\pi_u(\sigma\,|\,\epsilon,b)}{\pi_u(\sigma_0\,|\,\epsilon_0,b_0)} \;\propto\; \sqrt{\frac{\epsilon}{b+\epsilon\sigma}}$$

- Derive the conditional prior
- Caveat: **π(σ|n)** will be improper (unnormalizable) for a particular choice of Gamma priors unless one changes the nested sets.→
- Wish: A criterion to determine the nested set in a unique way

$$\pi(\sigma\,|\,n) \propto \int_0^\infty d\epsilon \int_0^\infty db\; \frac{(b+\epsilon\sigma)^n\,e^{-b-\epsilon\sigma}}{n!}\;\frac{\sqrt{\epsilon}\,\pi(\epsilon,b)}{\sqrt{b+\epsilon\sigma}}$$

L. Demortier

Reference Prior in Demand of a Code

- Analytical derivation of reference priors might be technically complicated

- Bernardo proposes an algorithm (pseudo code) to obtain a numerical approximation to the reference prior in the simple case of a one parameter model
  http://www.uv.es/~bernardo/RefAna.pdf
  The pseudo code should work for any number of parameters (of interest and nuisance)
  provided you make **non informative priors for ALL!**

- If code is extended to to multiple parameters, some including informative priors, it would even be more useful for the HEP community….

- The wish is to have a generalized routine (REAL CODE) to numerically calculate reference prior for parameters {θ} given the Likelihood L({θ}) as an input.

- Another complication is that the order of the parameters matter…..
  This should be further investigated and clarified!

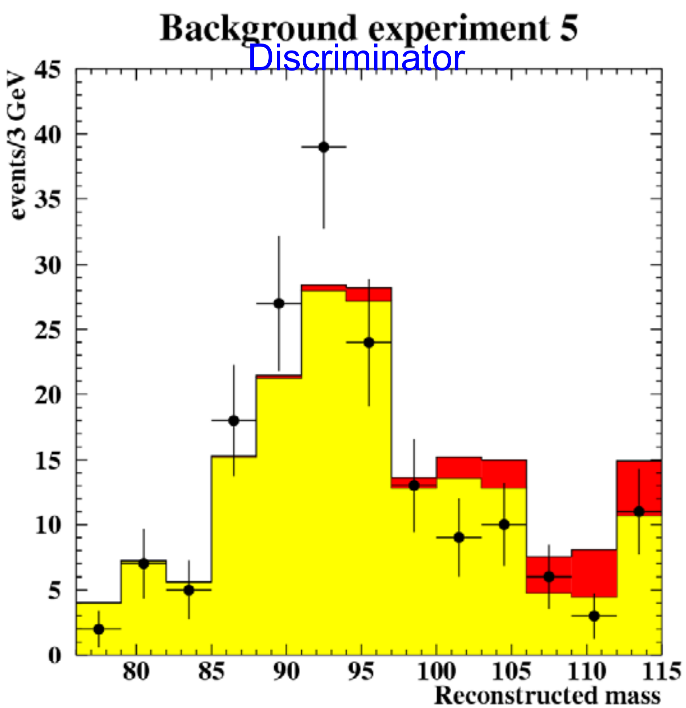
Frequentist
& Hybrid Methods

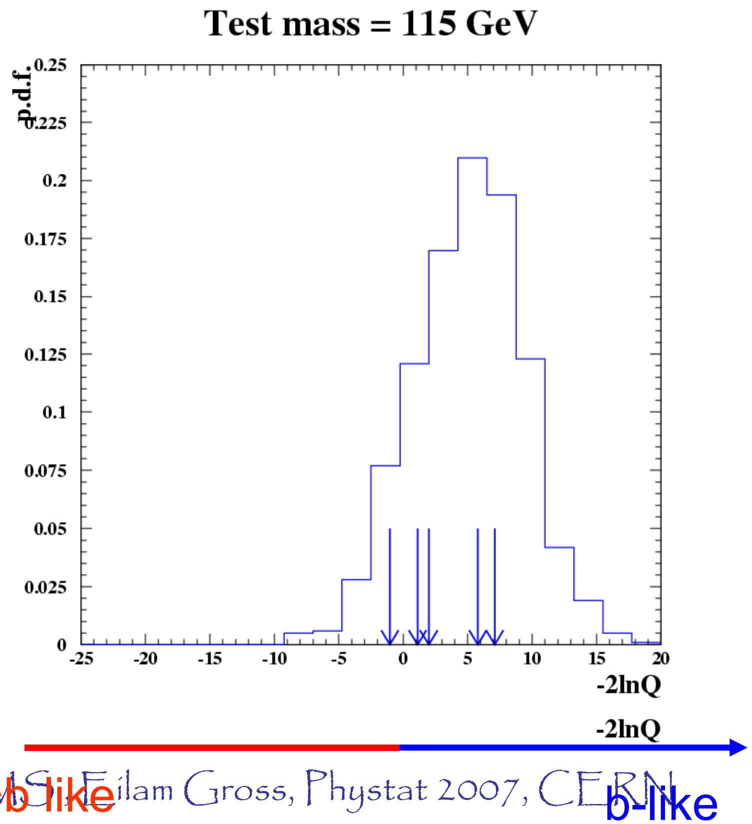
Example: Simulating BG Only Experiments

- The likelihood ratio, -2lnQ(m_H) tells us how much the outcome of an experiment is signal-like

$$Q(m) = \frac{L(H_1)}{L(H_0)} = \frac{L\left(s(m)+b\right)}{L(b)}$$

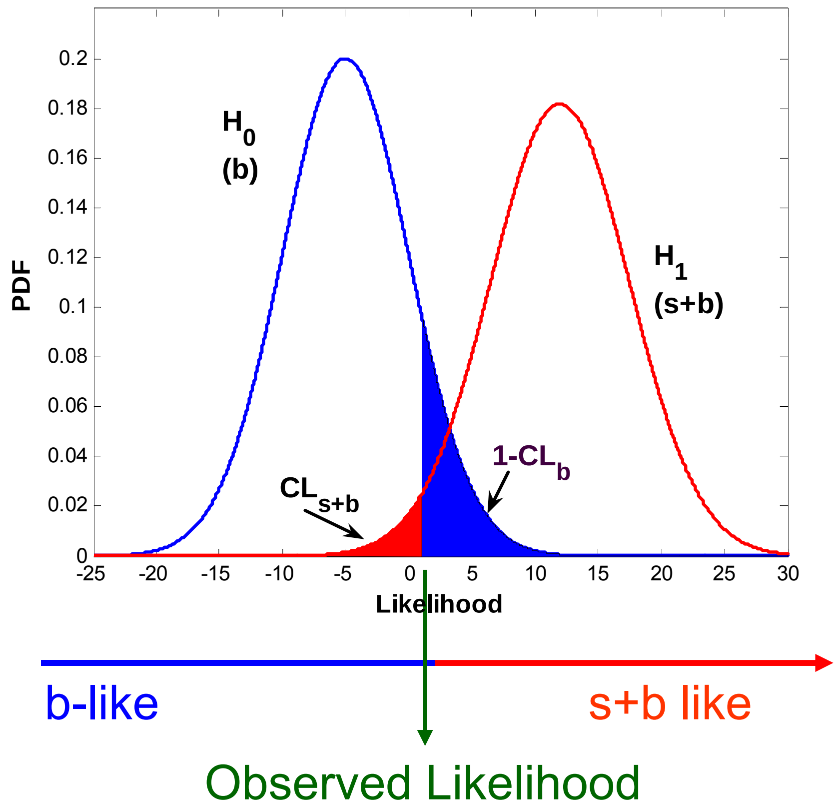
- NOTE, here the s+b pdf is plotted to the left!





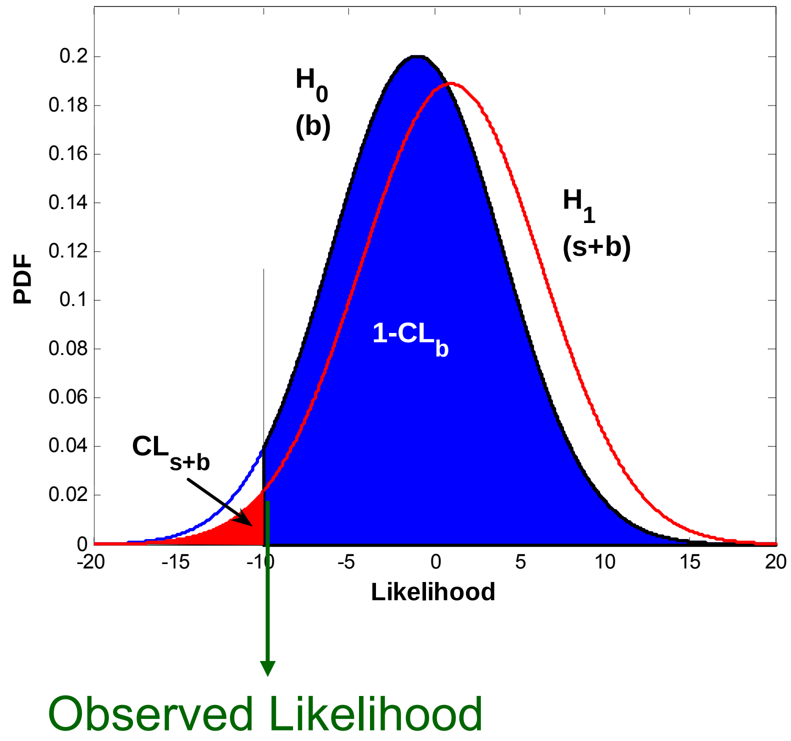
CL_{s+b} and CL_b

- $1\text{-}CL_b$ is the p value of the b-hypothesis, i.e. the probability to get a result less compatible with the BG only hypothesis than the observed one
(in experiments where BG only hypothesis is true)
- CL_{s+b} is the p-value of the s+b hypotesis, i.e. the probability to get a result which is less compatible with a Higgs signal when the signal hypothesis is true!
- **A small CL_{s+b} leads to an exclusion of the signal hypothesis at the 1- CL_{s+b} confidence level.**



The Problem of Small Signal

- $<N_{obs}>=s+b$ leads to the physical requirement that $N_{obs}>b$
- A very small expected s might lead to an anomaly when N_{obs} fluctuates far below the expected background, b.
- At one point DELPHI alone had $CL_{s+b}=0.03$ for $m_H=116$ GeV
- However, the cross section for 116 GeV Higgs at LEP was too small and Delphi actually had no sensitivity to observe it
- The frequntist would say: Suppose there is a 116 GeV Higgs….
  In 3% of the experiments the true signal would be rejected… (one would obtain a result incompatible or more so with m=116) i.e. a 116 GeV Higgs is excluded at the 97% CL…..



The CLs Method for Upper Limits

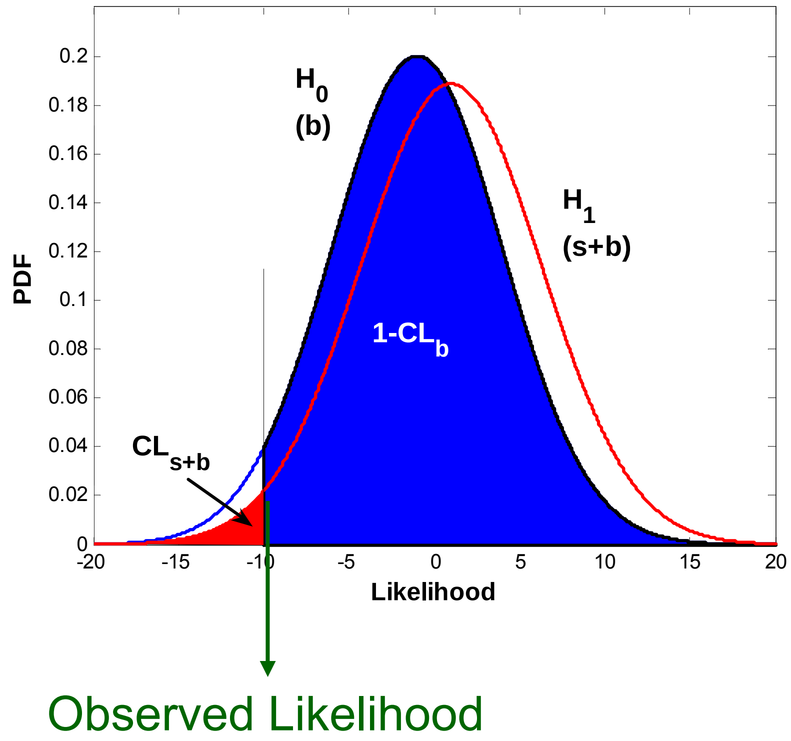
- Inspired by Zech(Roe and Woodroofe)'s derivation for counting experiments

$$P(n_{s+b} \le n_o | n_b \le n_o) = \frac{P(n_{s+b} \le n_o)}{P(n_b \le n_o)}$$

- A. Read suggested the CL_s method with

$$\boxed{CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1-p_b}}$$

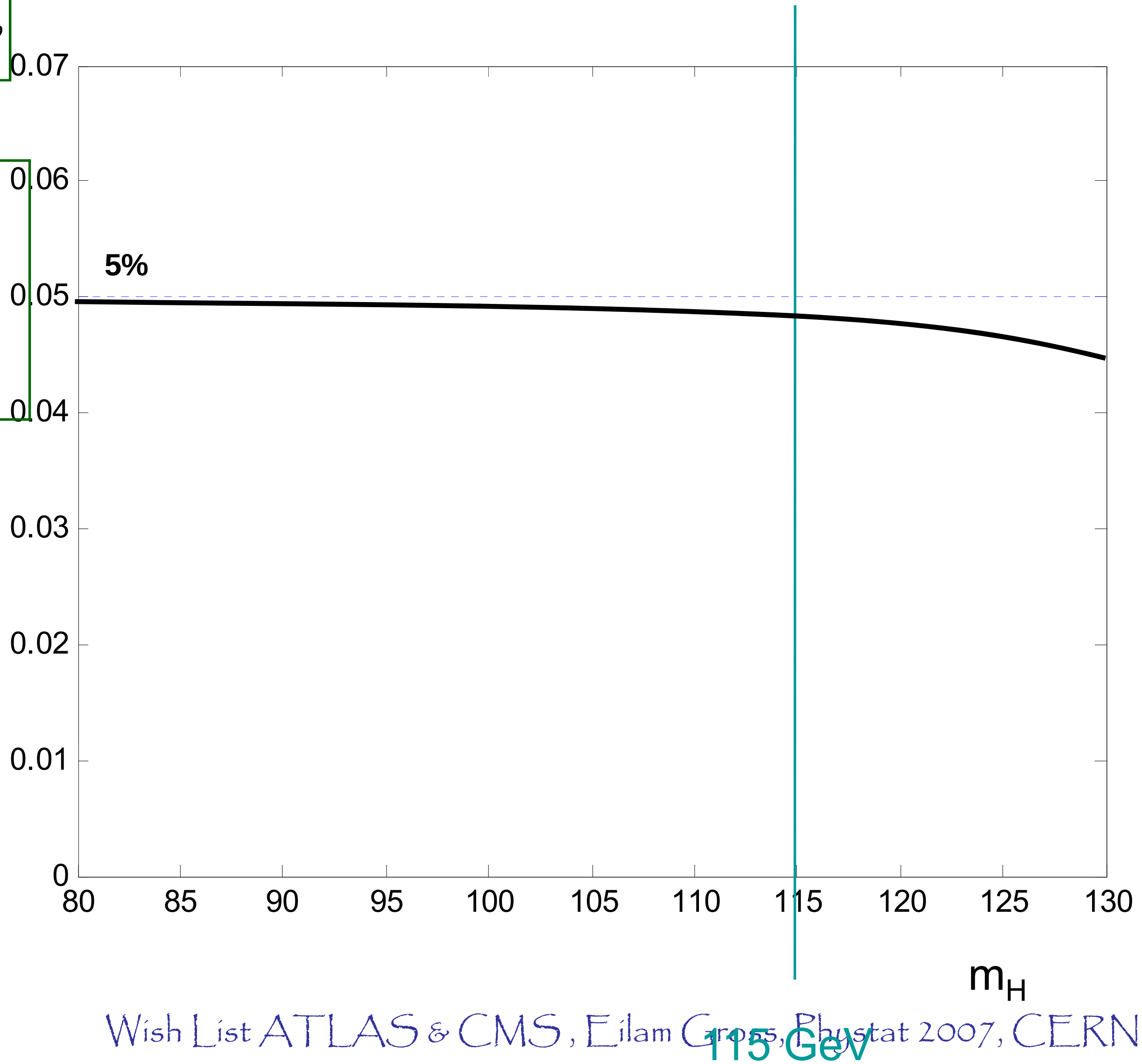
- In the DELPHI example, CLs=0.03/0.13=0.26, i.e. a 116 GeV could not be excluded at the 97% CL anymore….. (p_b=1-CL_b=0.87)



The Meaning of CLs

$$CL_s = \frac{CL_{s+b}}{CL_b} \xrightarrow{m_H \downarrow} CL_{s+b}$$

False exclusion rate of Signal when Signal is true



- Is it really that bad that a method undercovers where Physics is sort of handicapped… (due to loss of sensitivity)?

Basic Definitions

- Normally, we make **one** experiment and try to **estimate from this one experiment** the **confidence interval** at a specified CL% Confidence Level….

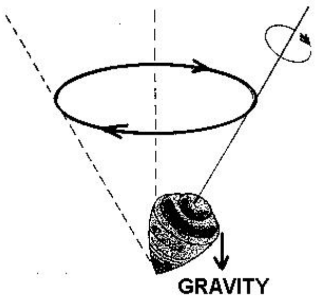
- In simple cases like Gaussians PDFs G(s,s_{true}) the Confidence Intrerval can be calculated analytically and ensures a complete coverage For example 68% coverage is precise for $\hat{s} \pm \sigma_{\hat{s}}$

Next.....

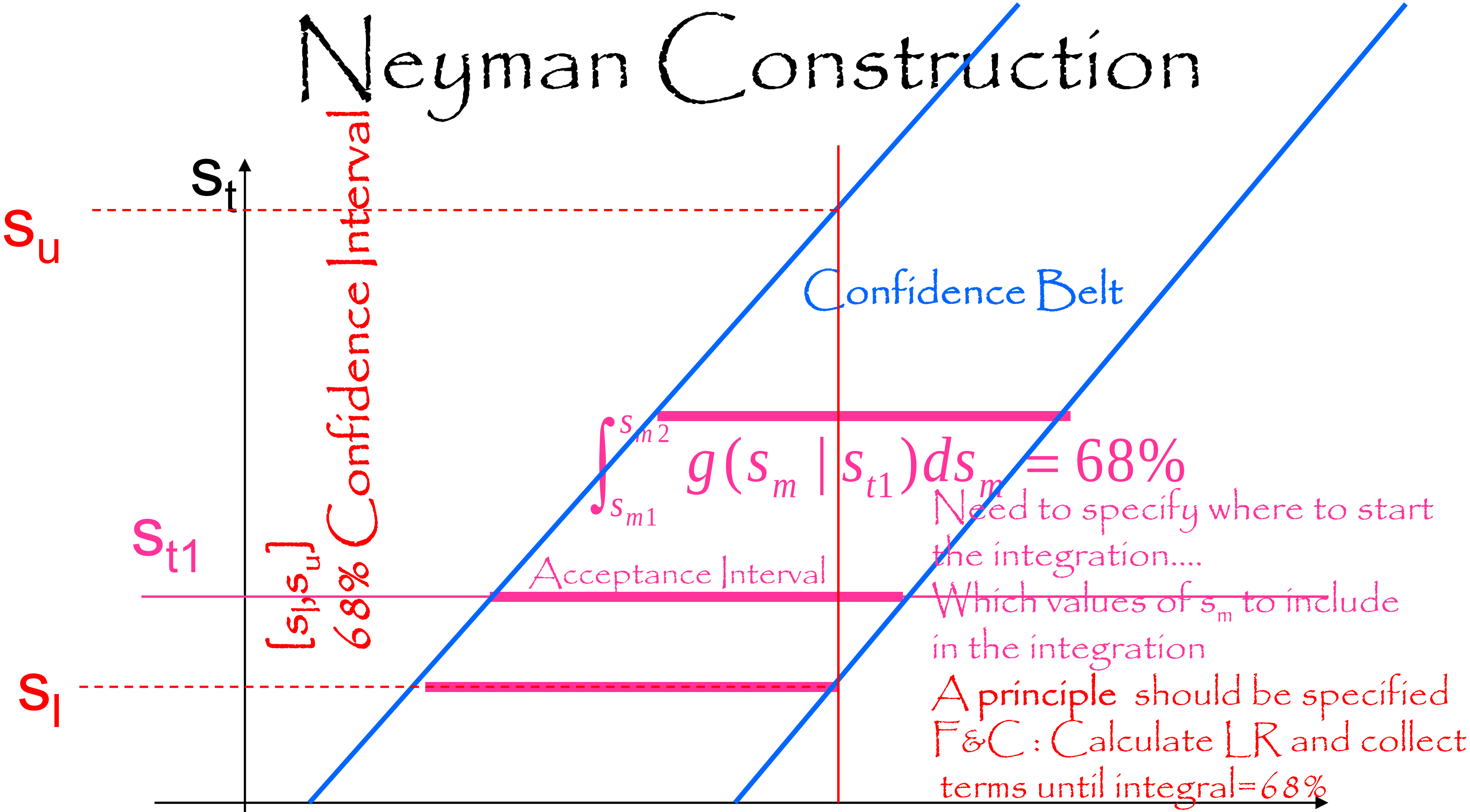
- F&C
- Profile likelihood – Full construction with nuisance parameters
- Profile construction
  (F&C construction with nuisance parameters)
- Profile - Likelihood

- For every complex problem, there is a solution that is simple, neat, and wrong. H.L. Mencken (quoted in Heinrich Phystat 2003)

- If you cannot explain it in a simple and neat way, you do not understand it
My brother…. (when, as an undergraduate student, I was trying to explain to him what is the meaning of precession…)



Neyman Construction



$[s_l, s_u]$ 68% Confidence Interval
In **68%** of the experiments the derived **C.I. contains** the **unknown true value of s**

- With Neyman Construction we guarantee a coverage via construction, i.e. for any value of the unknown true s, the Construction Confidence Interval will cover s with the correct rate.

The Flip Flop Way of an Experiment

- The most intuitive way to analyze the results of an experiment would be

  - Construct a test statistics
    e.g. $Q(x) \sim L(x|H_1) / L(x_{obs}|H_0)$

  - If the significance of the measured $Q(x_{obs})$, is less than 3 sigma, derive an upper limit (just looking at tables), if the result is >5 sigma (and some minimum number of events is observed….), derive a discovery central confidence interval for the measured parameter (cross section, mass….) …..

- **This Flip Flopping policy leads to undercoverage:**
  Is that really a problem for Physicists?
  Some physicists say, for each experiment quote always two results, an upper limit, and a (central?) discovery confidence interval

Frequentist Paradise – F&C Unified with Full Coverage

- Frequentist Paradise is certainly made up of an interpretation by constructing a confidence interval in brute force ensuring a coverage!
- This is the Neyman confidence interval adopted by F&C….

- The motivation:
  - Ensures Coverage
  - Avoid Flip-Flopping – an ordering rule determines the nature of the interval
    (1-sided or 2-sided depending on your observed data)
  - Ensures Physical Intervals

- Let the test statistics be $Q=L(s+b)/L(\hat{s}+b)=P(n|s,b)/P(n|\hat{s},b)$ where $\hat{s}$ is the
  physically allowed mean s that maximizes $L(\hat{s}+b)$
  (**protect a downward fluctuation of the background, $n_{obs}>b$**)

(Frequentist) Paradise Lost?

- A consequence of F&C ordering:
- An experiment with higher expected background which observes no events will set a better upper limit then an experiment with lower or no expected background
  - 95% upper limit for n_o=0,b=0 is 3.0
    95% upper limit for n_o=0, b=5 is 1.54
  - $P(n_{obs}=0|b=5)<P(n_{obs}=0|b=0)$
- Is the better designed analysis/experiment get punished?
- F&C claimit's a matter of education….
  The meaning of a confidence interval is NOT that given the data there is a 95% probability for a signal to be in the quoted interval…
  - NOT AT ALL… It means that given a signal, 95% of the possible outcome intervals will contain it But there are also 5% of possible intervals where the signal could be outside this interval
  - The experiment where the background fluctuated down from 5 to zero was lucky…. We probably fell in the 5% of the intervals where the signal could be above the quoted upper limit….(s_{true}>1.54)
    and the exclusion should have been weaker….
- **HOWEVER, if one repeats the experiment with no signal, one finds out that the average 95% CL is at 6.3 for b=5, i.e. the reported upper limit of 1.54 must have been sheer luck….**

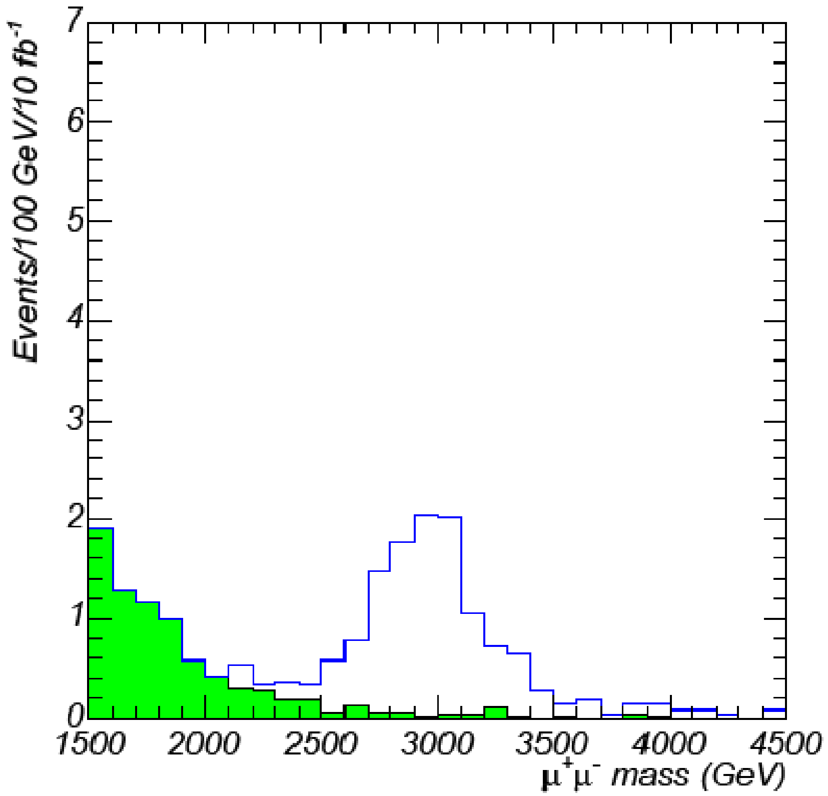
The Relevance to LHC

- A false claim: LHC deals only with 5 sigma discoveries
- With SUSY and Exotics you sometimes do not even know your signal…
- With Physics Beyond the SM we will continue to set limits!
  (Actually we will probably set many more limits than discoveries….)
- Another false claim: However, observing zero events is not something you will encounter… after a very short while….

Example: search for Z'

- Assume a heavy Z' (say 3 TeV)…
- For 10fb^{-1} L·σ·BR(Z'→mm)~20,

  L·σ_{DY}<2
- Why is this situation different from LEP?
  It is not!
- There always exist a Luminosity at which your signal is marginal
- At any given Luminosity there is always a Z' for which the signal is marginal….
  (thanks Bob)



CMS Note 2005/002
Cousins,Mumford,Valuev

F&C Relevance to LHC

- In the F&C method one does not have the freedom to choose the nature of the confidence interval. The fear is in those cases where one will get a 2-sided interval and exclude s=0 at the 95% CL… then what can one infer?

- Does this mean that you have made a discovery?

- Again, F&C will tell you it's a matter of education….
  The meaning of a Confidence interval which excludes 0, is not that **given the data** there is less than 5% probability for s=0….. It means that whatever the value of the true s is, it will not be included in 5% of the observed data derived intervals….. But perhaps your derived interval is exactly within the unlucky 5%?

- A legitimate confusion: So what can I infer from these intervals….
  ALWAYS check the average sensitivity of the experiment using BG only MC experiments!!!

- Is it possible to derive some quantitative expression that will take the expected sensitivity into account?
  (One can measure how many sigmas the obtained result is from the expected sensitivity….)

- CONFUSED? You won't be after the next episode…..

Including Systematics

C&H Hybrid Method

See example in Wade Fisher talk

- This method is coping with the Nuisance parameters by averaging on them weighted by a posterior.
- The Bayesian nature of the calculation is in the Nuisance parameters only….
- Say in a subsidiary measurement y of b, then the posterior is *p(b|y);* μ is the *x* expectation.
- C&H will calculate the p-value of the observation (x_o,y_o)

$$p(x_o,y_o\,|\,\mu)=\int_0^\infty p(x_o\,|\,y_o,\mu)p(b\,|\,y_o)db$$

$$p(b\,|\,y_o)=\frac{p(y_o\,|\,b)\,p(b)}{p(y_o)}$$

$$p(y_o\,|\,b)=G(y_o\,|\,b,\sigma_b)$$

$$p(b)\ \ uniform$$

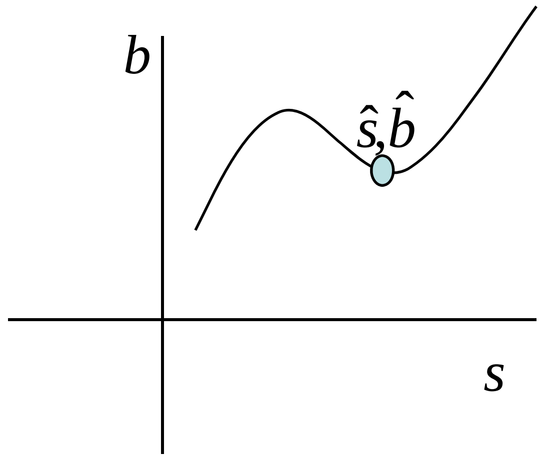
Note:
The original C&H used the Luminosity as the Nuisance parameter….

Full Neyman Constructions

Full Neyman Construction	
F&C w/b known	$LR(s)=\frac{L(s;b)}{L(\hat{s};b)}$

Full Neyman Constructions

Full Neyman Construction	
F&C w/b known	$LR(s) = \dfrac{L(s;b)}{L(\hat{s};b)}$
F&C w/b as a par of interest	$LR(s,b) = \dfrac{L(s,b)}{L(\hat{s},\hat{b})}$

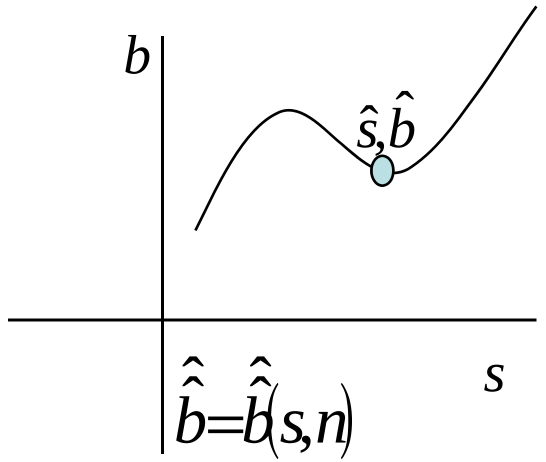


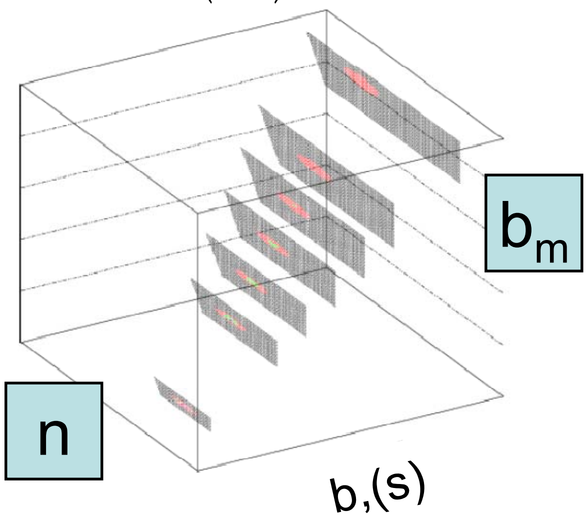
Full Neyman Constructions

Full Neyman Construction	
F&C w/b known	$LR(s) = \frac{L(s;b)}{L(\hat{s};b)}$
F&C w/b as a par of interest	$LR(s,b) = \frac{L(s,b)}{L(\hat{s},\hat{b})}$
Profile Likelihood FULL Construction	$\ell(s) = \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})}$

M. Kendall & A. Stuart

$$l = \frac{L(x \mid \theta_{r0}, \hat{\hat{\theta}}_s)}{L(x \mid \hat{\theta}_r, \hat{\theta}_s)}. \qquad (24.4)$$





An Extract about Full Neyman Construction For Both Signal and Nuisance Parameters

- I don't recommend you try this at home for the following reasons:
  - The ordering principle is not unique.
  - The technique is not feasible for more than a few nuisance parameters.
  - It is unnecessary since removing the nuisance parameters through profile likelihood works quite well.

    Gary Feldman, Concluding Remarks: Phystat 2005
- Is it really not feasible to do a FULL Neyman construction with >10 Nuisance parameters…..?
  - All we have seen so far with multi Nuisance parameters are semi-toy models….
- My preference is to eliminate at least the major nuisance parameters through profile likelihood and then do a (Profile) LR Neyman construction. It is straightforward and has excellent coverage properties.
  - **Gary Feldman**, Concluding Remarks: Phystat 2005

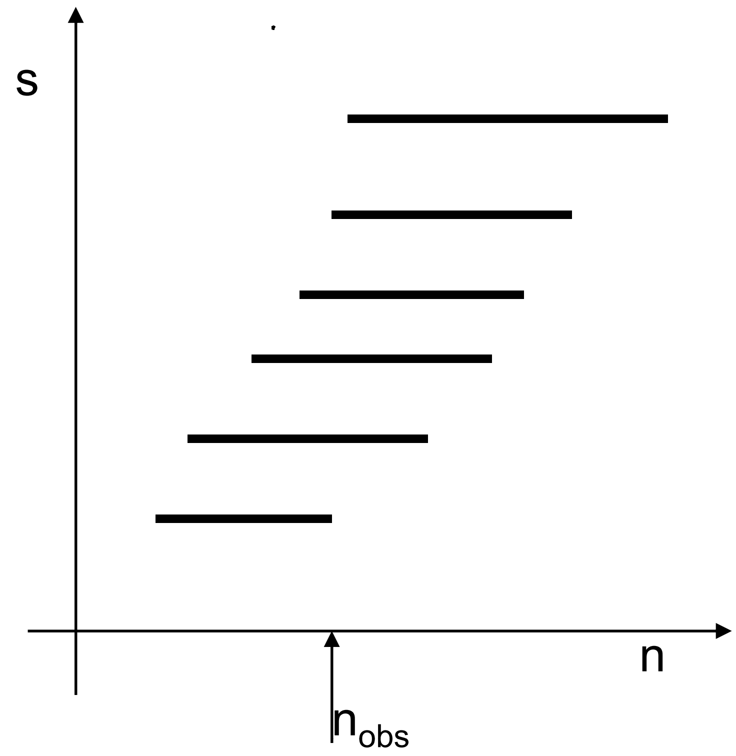
Approximate Neyman Contrsuctions

Profile Construction –
Construction with MLE

- Construct only at $\hat{\hat{b}} = \hat{\hat{b}}(s, n_{obs})$ with the order

$$\ell(s) = \frac{L(s, \hat{\hat{b}})}{L(\hat{s}, \hat{b})}$$

- Note: A more precised approximate construction could be with $\hat{\hat{b}} = \hat{\hat{b}}(s, n)$



The Profile Likelihood Method

$$\ell(s) = \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})} \Rightarrow Q(s) = \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})} \qquad -2\ln Q(s) = -2\ln \frac{L(s,\hat{\hat{b}})}{L(\hat{s},\hat{b})} \rightarrow \chi^2(s)$$

$$\Delta\chi^2 = 2.7 \rightarrow 90\%\ C.I.$$

- The advantages of the Profile Likelihood
  - It has been with us for years….. (MINOS of MINUIT) (**Fred James**)
  - In the asymptotic limit it is approaching a χ^2 distribution

F. James, e.g. Computer Phys. Comm. 20 (1980) 29 -35
W. Rolke, A. Lopez, J.Conrad. Nucl. Inst.Meth A 551 (2005) 493-503

The Recipes – Need to Choose One with Construction

	LH Principle	Coverage for the null (BG) hypothesis	Priors	Comments
F&C Neyman Construction without nuisance pars	No	Yes	No	**Exact**, proven to work but lack treatment of Nuisance parameters
F&C+ C&H Neyman Construction	No	No	Yes	
Profile Likelihood (Minuit)	Yes	NO (Asymptotic)	No	
F&C Profile Construction – Construction with MLE	No	Satisfactory	No	
Profile Likelihood – Full Construction	No	Yes	No	Impractical?
Bayesian	Yes	No	Yes	Choose Priors Carefully (reference, Jeffrey's)?
CLs with C&H (for Limits)	Yes		Yes	For Upper Limits ONLY!

Various Issues

L. Demortier

Looking for Guidelines on Subsequent Inference

- Fit a polynomial for background distribution
- Use a stepwise test to decide on the degree n
- The fitted coefficients a_i were obtained as if we know a priori the degree of the polynomial….
- How do you take the prior test into account? Perhaps the degree was wrong to start with?

$$\sum_{i=0}^{n} a_i x^i$$

Multivariate Analyses

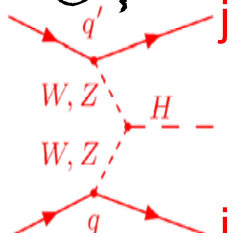
- The number of Physicists objecting to MV analyses (like ANN, Decision Trees) is getting smaller as the average year of birth of the active physicists go up…. (a personal prior….)

- Evaluating the systematics with MV analyses is very unclear…
  Many physicist have the habit of changing the input parameter by what they believe is a standard deviation… do it one at a time or randomly with all of them together…..

- There must be a better way to do it…

- Can the community come up with good figures of merit for the **robustness optimization of a MV analysis** (and not only for the significance $S/\sqrt{B}$)?

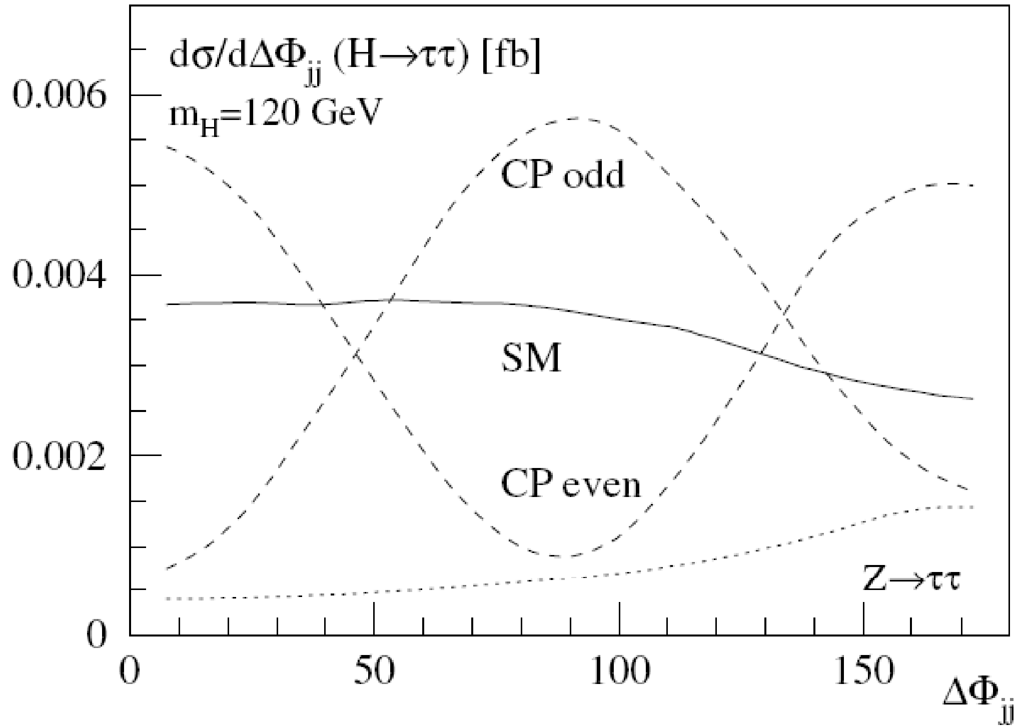
Telling Between Multi Hypotheses



- The Neyman-Pearson lemma tell us the best test statistics to tell between two simple hypotheses
- In case of more than one equivalent alternate hypotheses, what is the best test statistics to use besides testing them one against the other?

  (B. Cousins)

- Is there anyway to do it without a Bayesian assumption that all hypotheses have an a priory degree of belief?



D. Rainwater, D. Zeppenfeld, and K. Hagiwara, Phys. Rev. D **59**, 014037 (1999); T. Plehn, D. Rainwater, and D. Zeppenfeld, Phys. Lett. B **454**, 297 (1999); Phys. Rev. D **61**, 093005 (2000).

Fred James

Conclusions

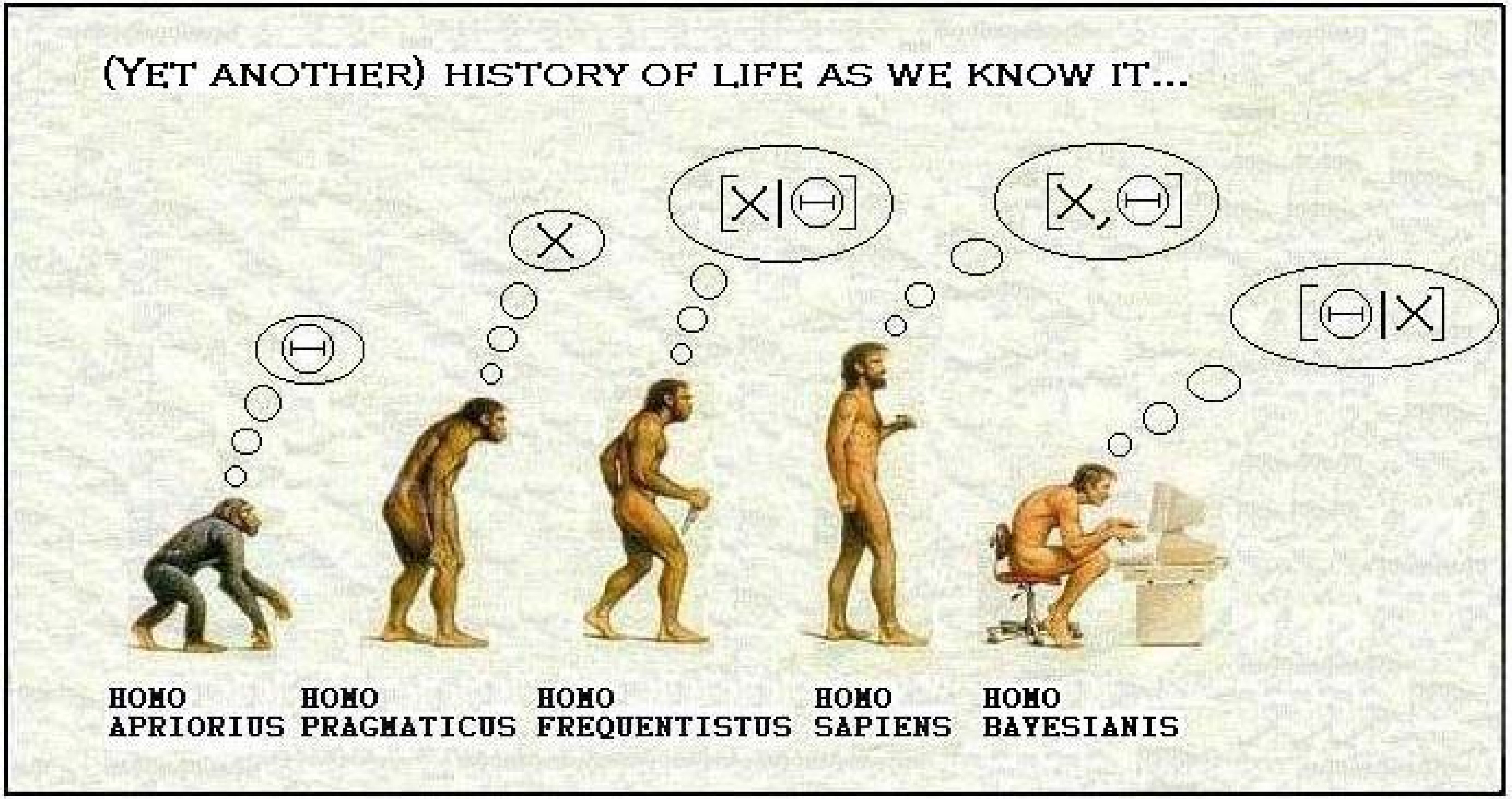
WHAT I WOULD LIKE TO SEE:

1. PHYSICISTS LEARN THE VOCABULARY OF STATISTICS
2. ASSUMPTIONS, METHODS, APPROXIMATIONS CLEARLY SPECIFIED IN PUBLICATIONS
3. FELDMAN/COUSINS IN ALL SEARCHES
4. BAYESIAN DECISION THEORY IN POLICY DECISIONS

- A partial Vocabulary:
  - p-value for discovery,
  - Significance,
  - CL, CI,
  - Neyman Construction,
  - Profile Likelihood,
  - Priors
- Explain your method, your priors
- Recipe (all with systematics):
  - Full or approximate Neyman Construction
  - Profile Likelihood
  - Bayesian
  - CLs for upper limits

A personal wish: Educate every physicist to become a Phystatistician to some degree

The End


(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[Θ]
[X]
[X|Θ]
[X,Θ]
[Θ|X]
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS

Backup Transparencies & Extensions

Fred James
In the first Workshops on Confidence Limits
CERN & Fermilab, 2000

- Many physicists will argue that Bayesian methods with informative physical priors are very useful

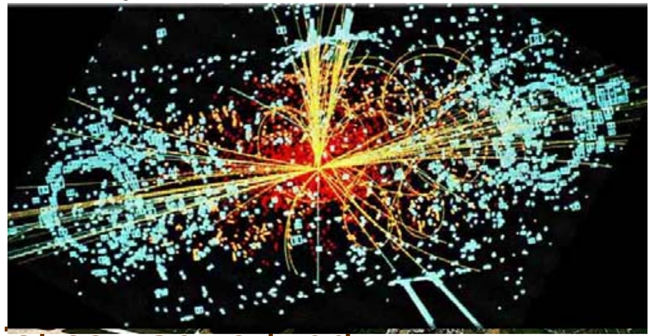
SUMMARY

FREQUENTIST METHODS STILL OFFER THE ONLY WAY TO PRESENT EXPERIMENTAL RESULTS OBJECTIVELY WITH THE USUAL SCIENTIFIC MEANING.

BUT

- BAYESIAN METHODS ARE GOOD FOR DECISION MAKING.
  DO PHYSICISTS MAKE DECISIONS?
- BAYESIAN METHODS ARE GOOD FOR BETTING
  DO PHYSICISTS MAKE BETS?
- BAYESIAN METHODS ARE GOOD WHEN THERE IS A PRIOR PROBABILITY OR PHASE SPACE
  MAXIMUM ENTROPY METHOD
- BAYESIAN METHODS ARE A GOOD WAY TO COMBINE NEW KNOWLEDGE WITH PRIOR BELIEFS.
  DO WE DO THIS?

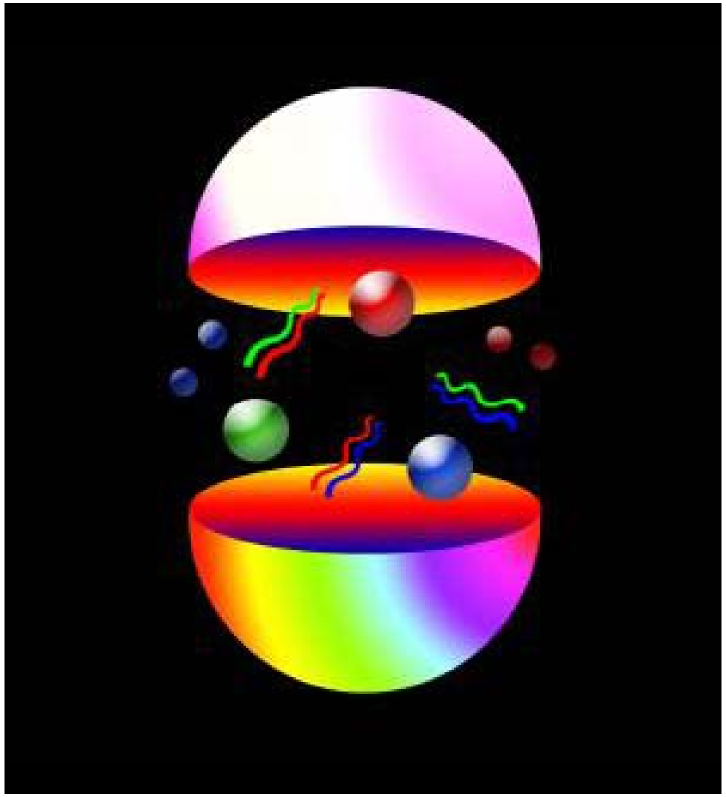
Modeling of the Underlying Processes



- Proton-Proton collisions
- Understanding PDFs or pdfs
  - Parton Density Functions……
  - This is our first encounter of the statistics vague kind….
- Hard Processes (LO,NLO,NNLO…)
- Herwig vs Pythia vs …..:
  The main difference between PYTHIA and HERWIG (e.g.) is the hadronisation model: How Quarks and Gluons manifest themselves into jets of particles….. Both models are in a way “equally correct” from the theoretical/experimental point of view… none is carved in stone…..

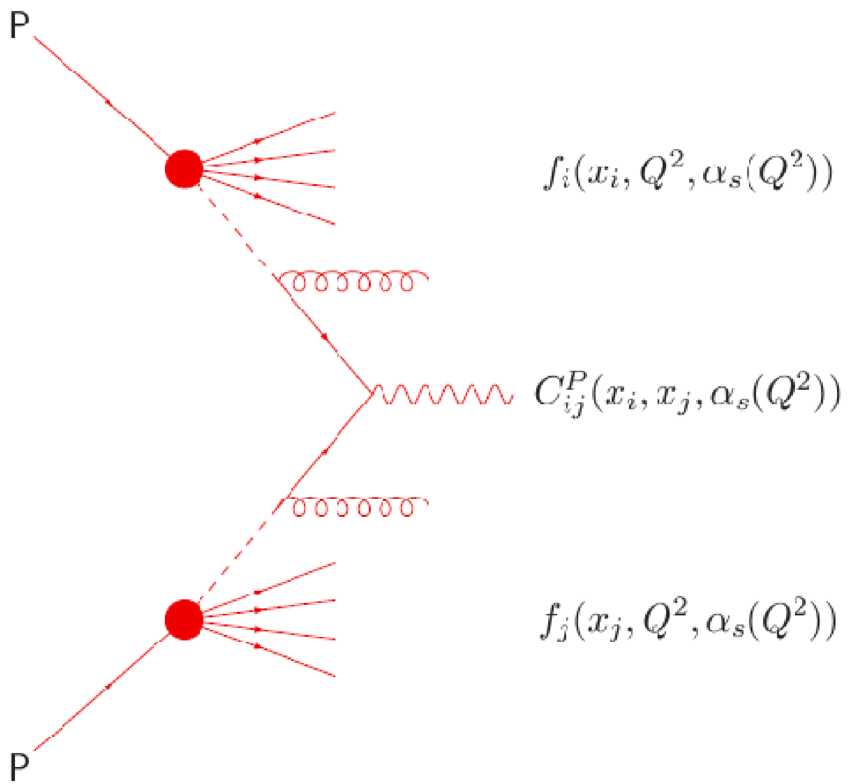
Modeling of the Underlying Processes

- Partons – Quarks and Gluons, the constituents of the Proton
- The Proton is made of the u,u,d which are the valence Quarks which interact via the exchange of gluons that can generate a sea of all sorts of quarks, anti-quarks and gluons
- The probability density to find in the proton a parton i, which carries a fraction x of the Proton momentum is dependent on the energy scale Q^2 and is given by $f_i(x;Q^2,\alpha_s(Q^2))$



pdf

$$\sigma(pp \to X) \sim x_i f_{i(x_i, Q^2, \alpha_s(Q^2))} \otimes \sigma(ij \to X)_{(x_i, x_j, \alpha_s(Q^2))} \otimes x_j f_{j(x_j, Q^2, \alpha_s(Q^2))}$$



Robert Thorne
(See TALK)

- Note that the structure functions are NOT observables!
- What we measure are e.g. cross sections of various processes from which we try to infer the structure functions and the strong couplings

$$\alpha_s(Q^2) \approx \frac{4\pi}{(11 - 2/3N_f)\ln(Q^2/\Lambda^2_{QCD})}$$

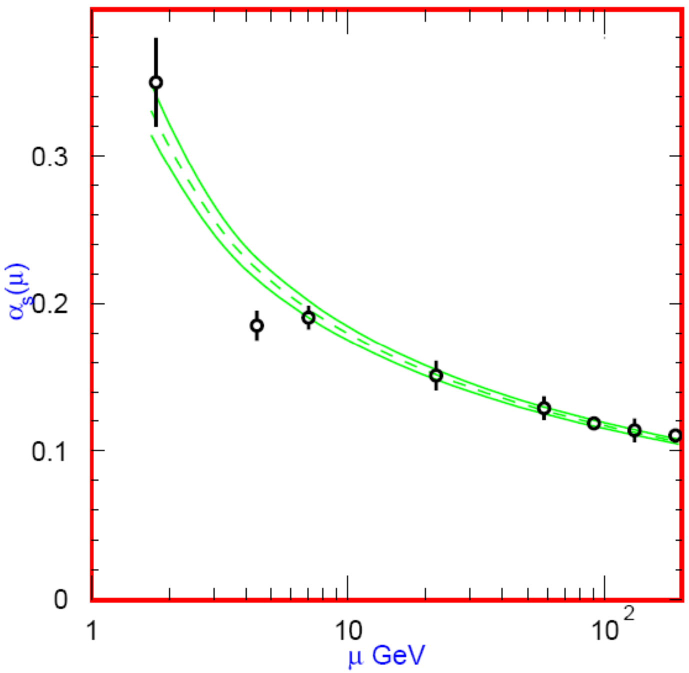
α_s is the strong coupling constant and is valid for Q^2>2 GeV (where perturbation holds)

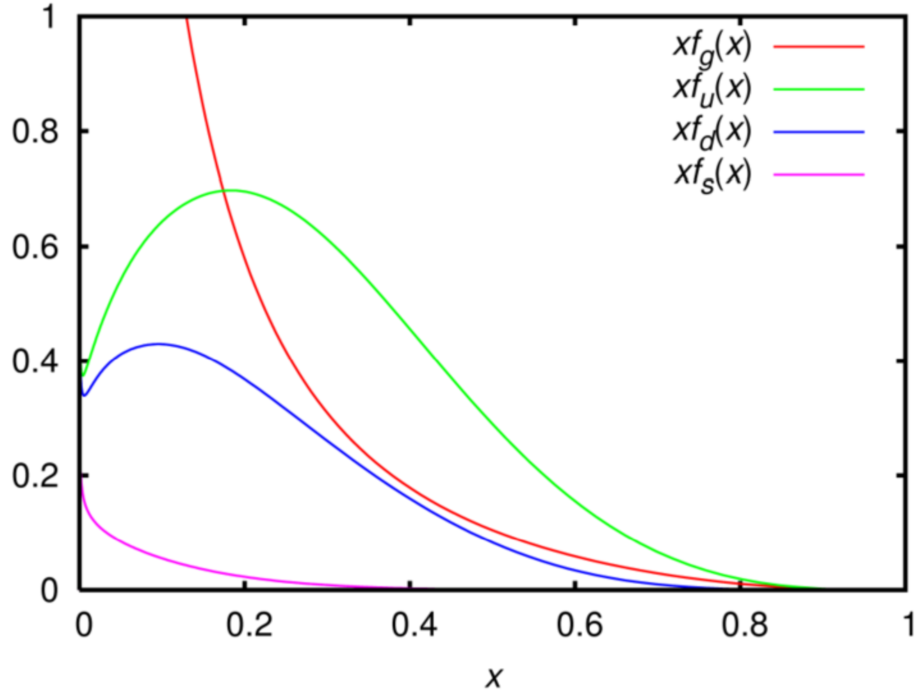
pdf: Parton Distribution Functions

- The revenge of the Physicist:
- פונקציות ההתפלגות הפרטוניות מהוות את צפיפות ההסתברות למציאת חלקיק עם תנע...

- The Parton Distribution Functions are the probability density for finding a particle with a certain longitudinal momentum fraction *x* and momentum transfer within an hadron.

$$f_i(x; Q^2, \alpha_s(Q^2))$$

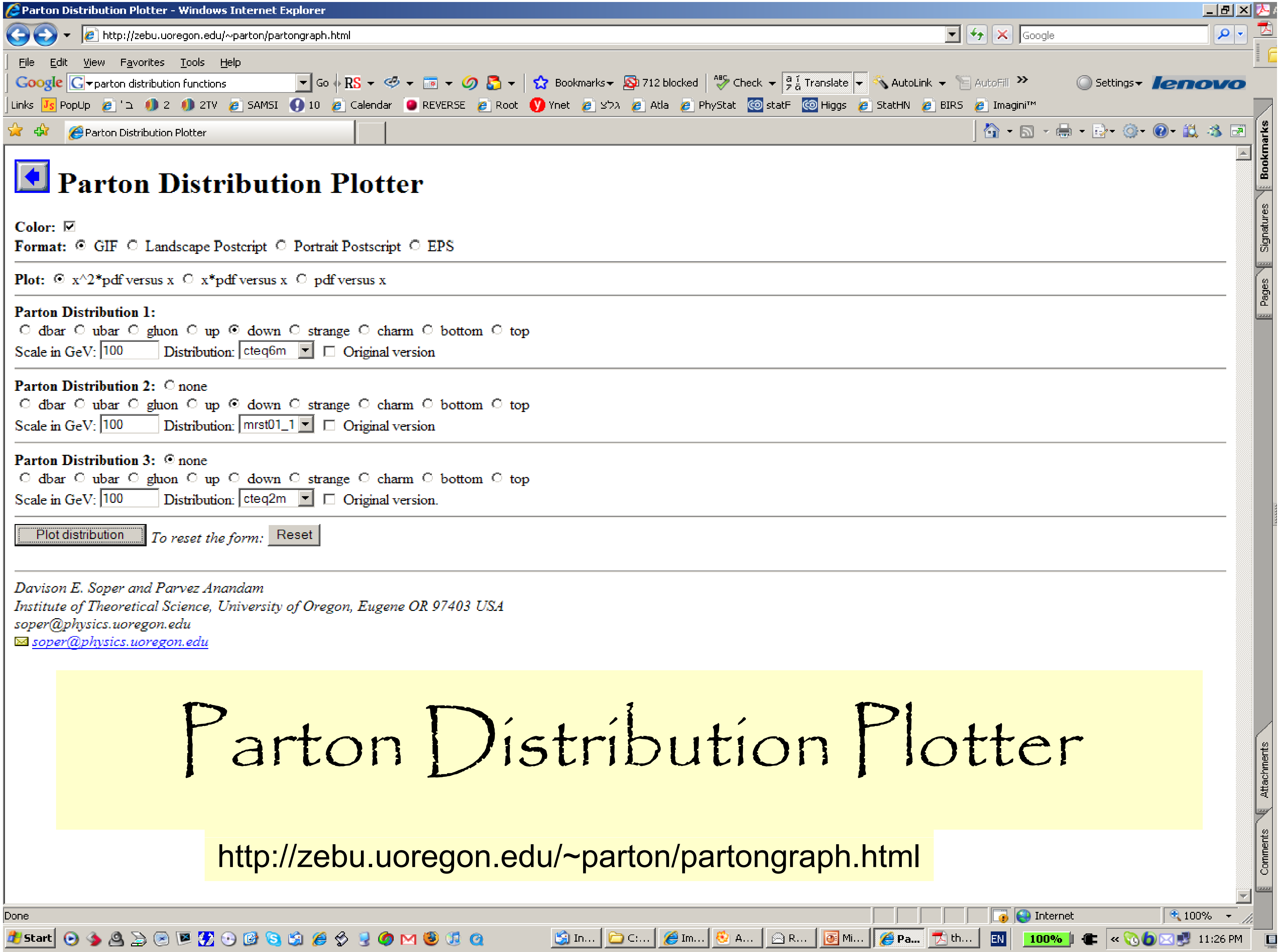




$$\sigma(pp \to X) \sim x_i f_{i(x_i,Q^2,\alpha_s(Q^2))} \otimes \sigma(ij \to X)_{(x_i,x_j,\alpha_s(Q^2))} \otimes x_j f_{j(x_j,Q^2,\alpha_s(Q^2))}$$

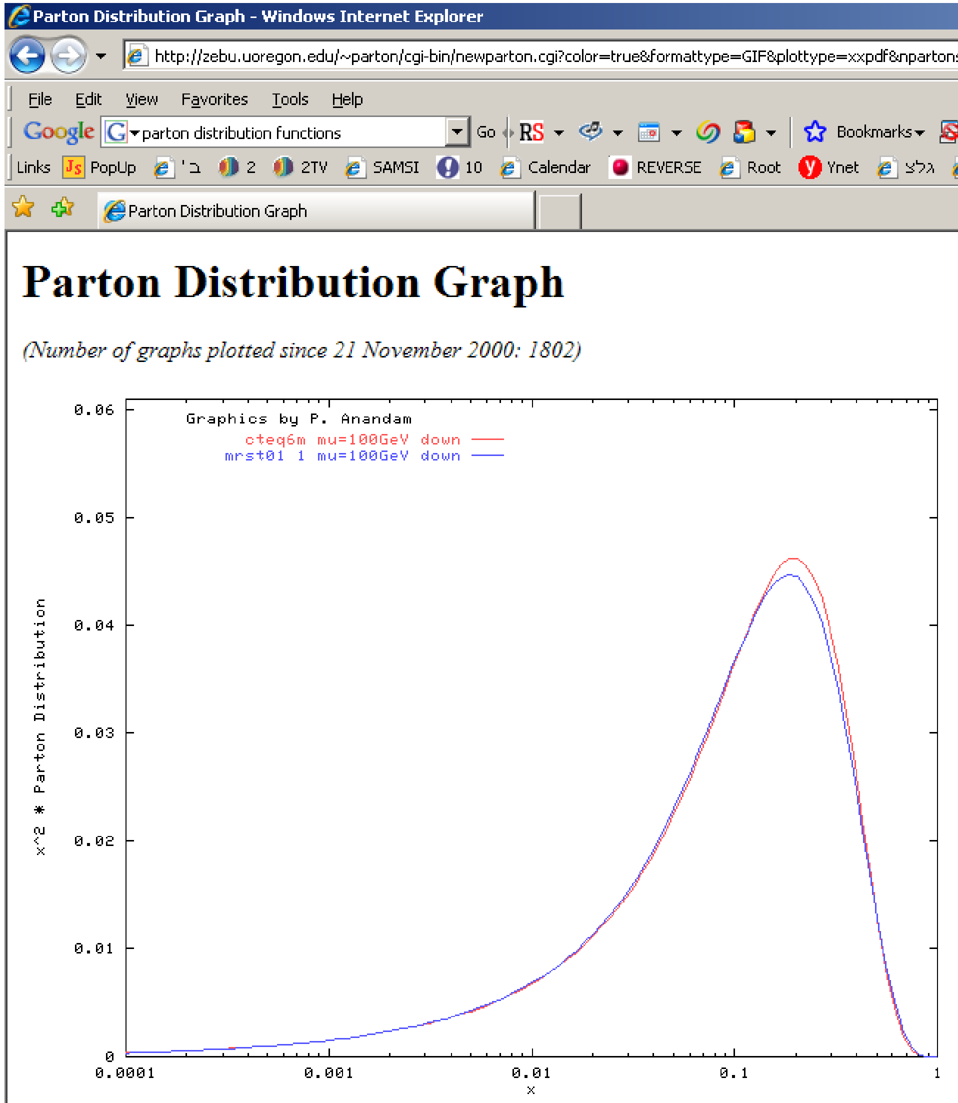
α_s is the strong coupling constant and
is valid for Q^2>2 GeV (where perturbation holds)

$$\alpha_s(Q^2) \approx \frac{4\pi}{(11 - 2/3N_f)\ln(Q^2/\Lambda^2_{QCD})}$$



Parton Distribution Plotter

http://zebu.uoregon.edu/~parton/partongraph.html



- The fitted pdf and couplings are then used to predict new unknown processes…

$$\boxed{\sigma(pp \to H) \sim x_i f_{i(x_i, Q^2, \alpha_s(Q^2))} \otimes \sigma(ij \to H)_{(x_i, x_j, \alpha_s(Q^2))} \otimes x_j f_{j(x_j, Q^2, \alpha_s(Q^2))}}$$

- If two data sets do not agree with each other, no theoretical model can agree with them
- To give weights to different data sets (according to which seems more consistent and better) is not acceptable
- What's happening?
  - Must be systematic errors, experimental or theoretical, that work against the statistics common sense….

The “Excuse”

- From the introduction to D. Stump talk in Phystat 2003:
  – The program of global analysis is not a routine statistical calculation, because of systematic errors—both experimental and theoretical. Therefore we must sometimes use physics judgment in producing the PDF model, as an aid to the objective fitting procedures.

- …Therefore we must sometimes use priors in a Bayesian manner in order (e.g.) to regulate the smoothness of the pdf.

Interpreting The Results

- Significance
- Hypothesis testing
- Limits or Discovery

  - So there are two alternate questions
    (Do not confuse between them):
    - Did I or did I not establish a discovery
      - Goodness of fit (get a p-value based on LR)
        What value of control sample should I pre-define for discovery?
        3σ equivalent? 5σ equivalent? Is the 5σ a myth?
    - How well my alternate model describes this discovery (measurement…., here one is interested in a confidence interval [m_l,m_u], one aims at a statement about the true value of the measured parameter, say, the Higgs mass , the probability of the Higgs to be in the mass range $m_H \in$ [m_l,m_u] is 99%....
      IS IT?
    - No it isn't….., that's not the right answer, or did I pose the right question? (more later)

Blind Analysis
"2c or not 2c"

- Blind analysis is not really an issue of Folded or Unfolded ("2c or not 2c")… but an issue of being objective in the analysis and release the collaboration from the need to deal with spurious peaks every other day…

- However, since it seems that at least in the first few year(s) there is a semi consensus that this question is irrelevant, and nobody is going to fold any eyes…. I will not discuss it here or present any questions….

Basic Definitions: Likelihood Principle

- The likelihood principle as stated by Edwards (1972):
  - Within the framework of a statistical model, *all* the information which the data provide concerning the relative merits of two hypotheses is contained in the likelihood ratio of those hypotheses on the data.
  - For a continuum of hypotheses, this principle asserts that **the likelihood function contains all the necessary information.**
- **The LH Principle is a corollary of Bayesian statistics**
- **Fisher (1932):** "...when a *sufficient* statistic exists, that is one which in itself supplies the whole of the information contained in the sample…."
- This is one reason, why asymptotically many methods, like the MINOS/MINUIT (Profile Likelihood), work so well, because asymptotically they fulfill the Likelihood Principle….
- So you do not need to be a Bayesian if you insist on the Likelihood Principle….

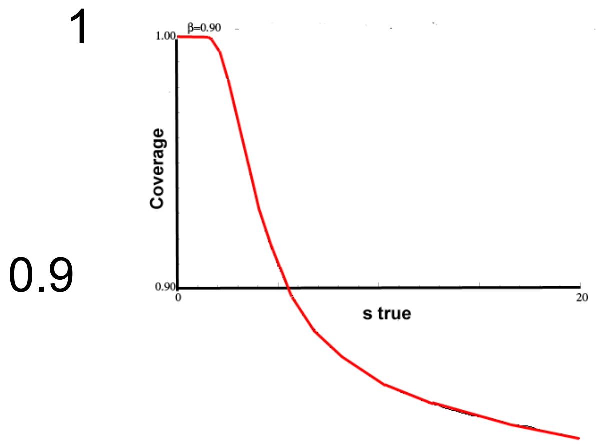
Nuisance Parameters (Systematics)

- Nuisance – a thing causing inconvenience or annoyance (Oxford Dictionary)
- Systematic Errors are equivalent in the statisticians jargon to Nuisance parameters – parameters of no interest…
  Will the Physicist ever get used to this jargon?
- D. Sinervo classified uncertainties into three classes (see Kyle's talk):
  - **Class I**: Statistics like – uncertainties that are reduced with increasing statistics. Example: Calibration constants for a detector whose precision of (auxiliary) measurement is statistics limited
  - **Class II**: Systematic uncertainties that arise from one's limited knowledge of some data features and cannot be constrained by auxiliary measurements … One has to do some assumptions. Example: Background uncertainties due to fakes, isolation criteria in QCD events, shape uncertainties…. These uncertainties do not normally scale down with increasing statistics
  - **Class III**: The "Bayesian" kind… The theoretically motivated ones… Uncertainties in the model, Parton Distribution Functions, Hadronization Models…..

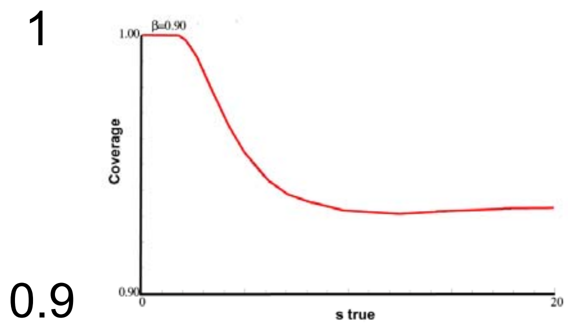
Bayesian Priors

4-channel Poisson with signal aceptance and background nuisance parameters

Flat priors



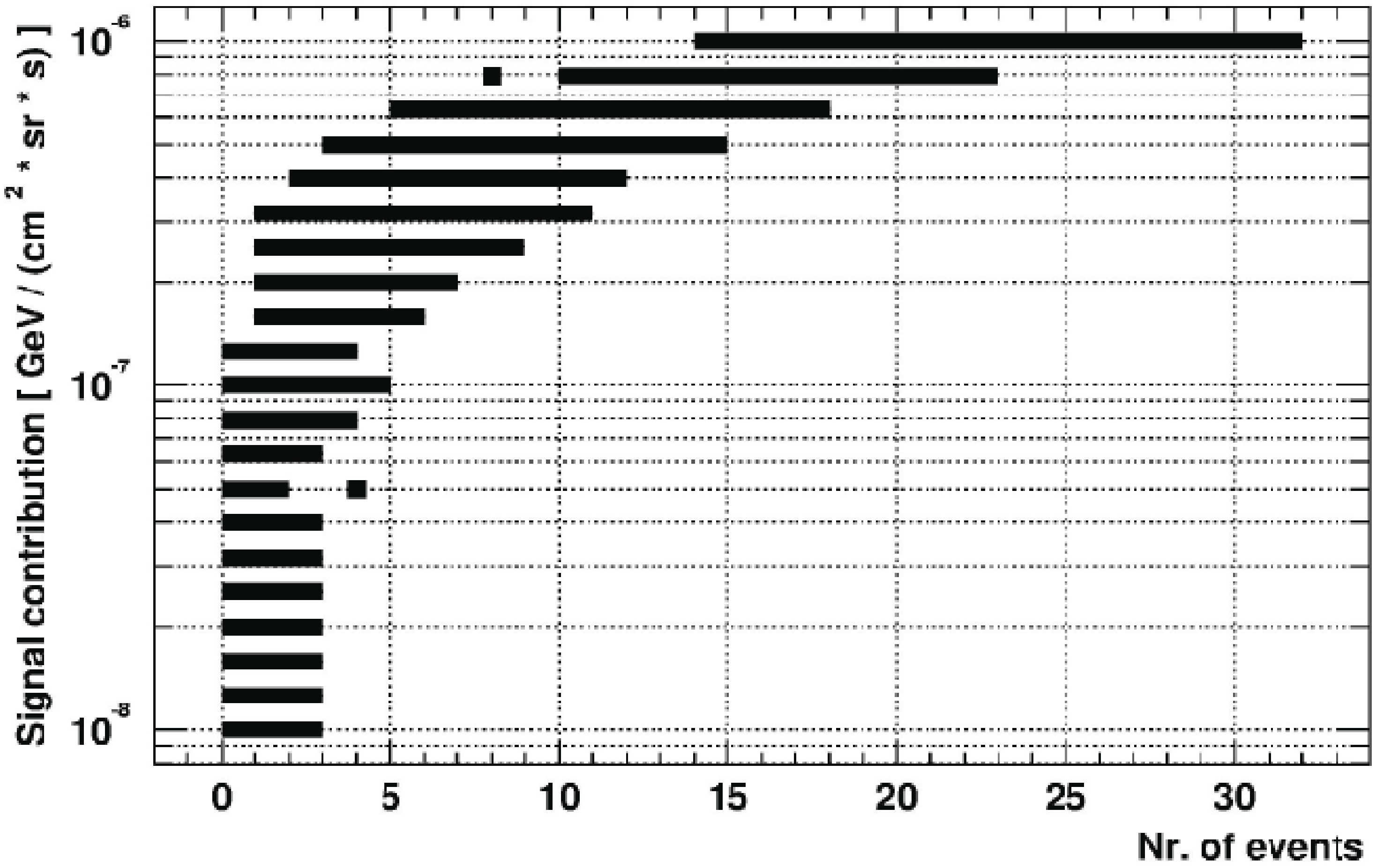
1/ε and 1/*b* priors



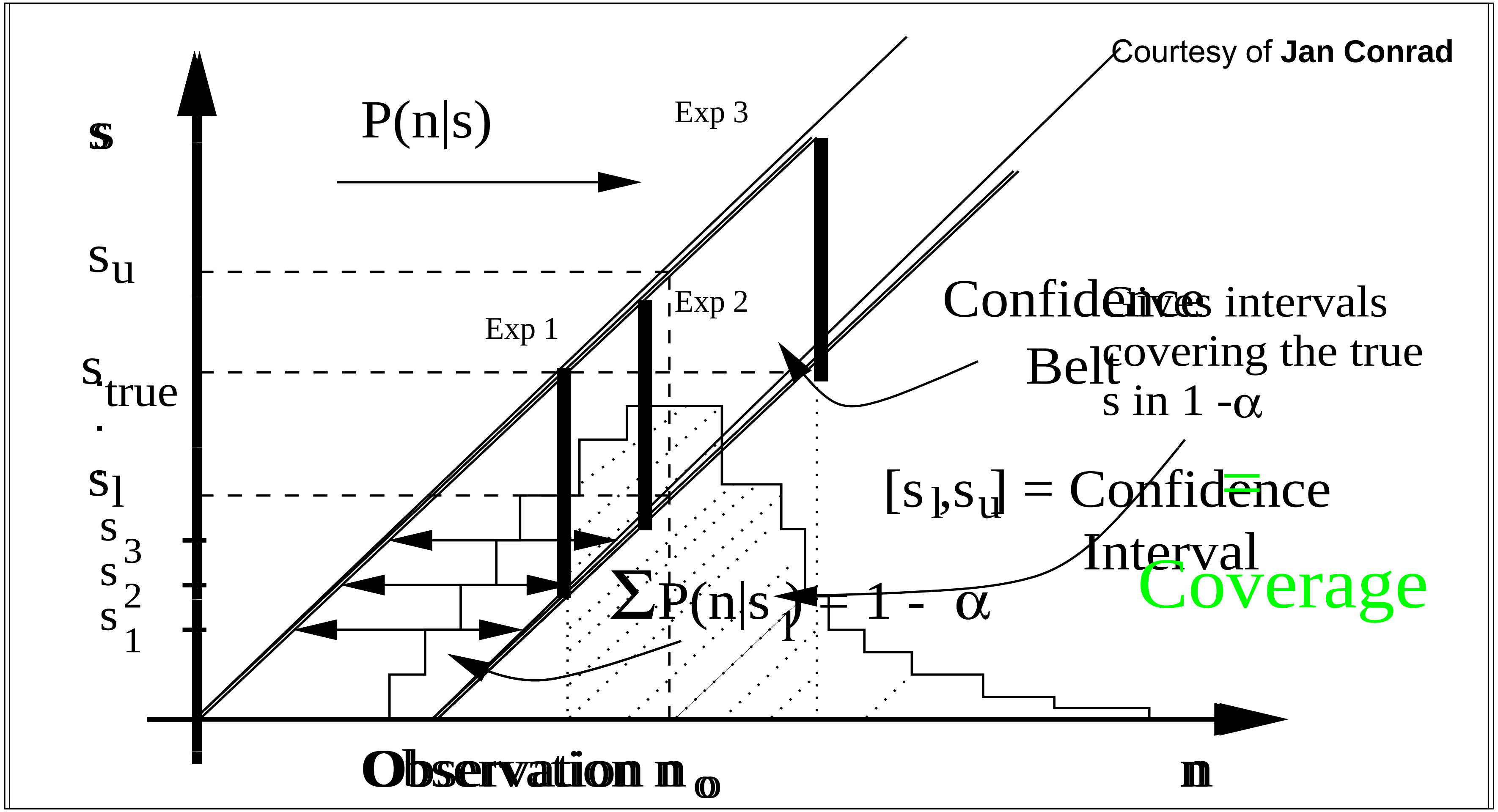
–Joel Heinrich Phystat 2005

– Note that the use of Bayesian priors with frequentist like coverage might be more work then performing a full Neyman construction

Example: Non Trivial Neyman Belt



Basic Definitions: A Neyman Construction



- With Neyman Construction we guarantee a coverage via construction, i.e. for any value of the unknown true s, the Construction Confidence Interval will cover s with the correct rate.

The Flip Flop Way of an Experiment

- The most intuitive way to analyze the results of an experiment would be
  - Construct a test statistics
    e.g. $Q(x) \sim L(x|H_1)/ L(x_{obs}|H_0)$
  - If the significance of the measured $Q(x_{obs})$, is less than 3 sigma, derive an upper limit (just looking at tables), if the result is >5 sigma (and some minimum number of events is observed….), derive a discovery central confidence interval for the measured parameter (cross section, mass….) …..
- **This Flip Flopping policy leads to undercoverage:**
  If we construct an a priori 90% Confidence acceptance interval [x1,x2] it turns out that there are true values of x such that $P(x_{true} \in [x1,x2]) < 90\%$
- Is that really a problem for Physicists?
  Some physicists say, for each experiment quote always two results, an upper limit, and a (central?) discovery confidence interval

Frequentist Paradise – F&C Unified with Full Coverage

- Frequentist Paradise is certainly made up of an interpretation by constructing a confidence interval in brute force ensuring a coverage!
- This is the Neyman confidence interval adopted by F&C….

- The motivation:
  - Ensures Coverage
  - Avoid Flip-Flopping – an ordering rule determines the nature of the interval (1-sided or 2-sided depending on your observed data)
  - Ensures Physical Intervals

- Let the test statistics be Q=L(s+b)/L(ŝ+b)=P(n|s,b)/P(n|ŝ,b) where ŝ is the **physically allowed** mean s that maximizes L(ŝ+b)
  (**protect a downward fluctuation of the background, n_{obs}>b**)
  Here you already see the power of F&C to tell between two possible signals separated well from the background.

- To ensure coverage use a Neyman reconstruction with the ordering determined by Q, i.e. construct [n1,n2] for s, such that

$$\sum_{n1}^{n2} p(n\,|\,s,b) \geq CL, \forall\ l \notin [n1,n2], n \in [n1,n2], Q(l) < Q(n)$$

- The confidence interval (in the true parameters space) will be constructed using the confidence belt, once the experiment is performed and n_o events are observed.

(Frequentist) Paradise Lost?

- A consequence of F&C ordering:
- An experiment with higher expected background which observes no events will set a better upper limit then an experiment with lower or no expected background
  - 95% upper limit for n_o=0,b=0 is 3.0
    95% upper limit for n_o=0, b=5 is 1.54
  - $P(n_{obs}=0|b=5)<P(n_{obs}=0|b=0)$
- Is the better designed analysis/experiment get punished?
- F&C claimit's a matter of education….
  The meaning of a confidence interval is NOT that given the data there is a 95% probability for a signal to be in the quoted interval…
  - NOT AT ALL… It means that given a signal, 95% of the possible outcome intervals will contain it But there are also 5% of possible intervals where the signal could be outside this interval
  - The experiment where the background fluctuated down from 5 to zero was lucky…. We probably fell in the 5% of the intervals where the signal could be above the quoted upper limit….(s_{true}>1.54) and the exclusion should have been weaker….
- **HOWEVER, if one repeats the experiment with no signal, one finds out that the average 95% CL is at 6.3 for b=5, i.e. the reported upper limit of 1.54 must have been sheer luck….**

Using the General C&H

- Fold PDF (for prime process) with a PDF describing the uncertainties, e.g.

$$q(n)_{s+b} = \frac{1}{2\pi\sigma_b\sigma_\epsilon}\times \int_0^\infty \int_0^\infty p(n)_{b'+\epsilon' s}\, e^{\frac{-(b-b')^2}{2\sigma_b^2}}\, e^{\frac{-(1-\epsilon')^2}{2\sigma_\epsilon^2}}\, db'\, d\epsilon'$$

J. Conrad

- Next step, perform a Neyman reconstruction with the folded PDF and F&C ordering

The F&C Profile

$$l(x,\theta_{r_0}) = \frac{L(x\,|\,\theta_{r_0},\hat{\hat{\theta}}_s)}{L(x\,|\,\hat{\theta}_r,\hat{\theta}_s)}.$$

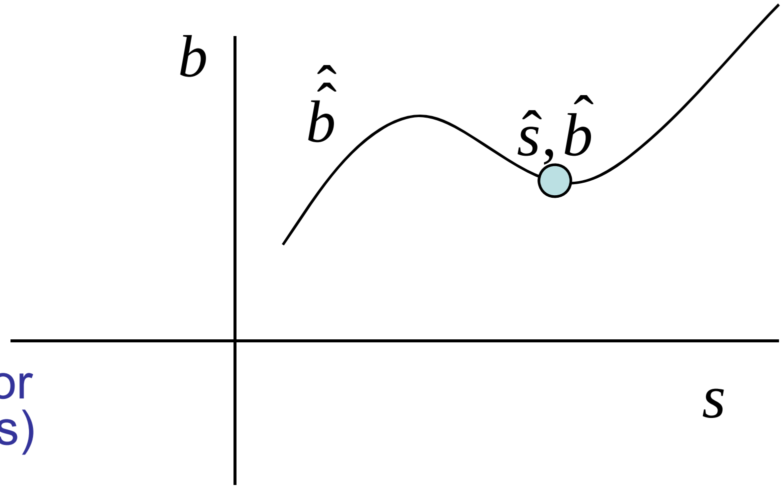
- Gary Feldman in his talk in the FermiLab workshop starts from describing the chapter in Kendall and Stuart dealing with the elimination of the nuisance parameters by maximizing the likelihood with respect to them

$$\boxed{l(n,s) = \frac{L\left(n,b_m\,|\,s,\hat{\hat{b}}\right)}{L\left(n,b_m\,|\,\hat{s},\hat{b}\right)}}$$

here, b_m is an auxiliary measurement of b.

$\hat{\hat{b}}$ Is the favorable value of b for s…

$\hat{s},\hat{b}$ MLE



- Now one can use this LR as an ordering for a Neyman construction in the space of (n,s)

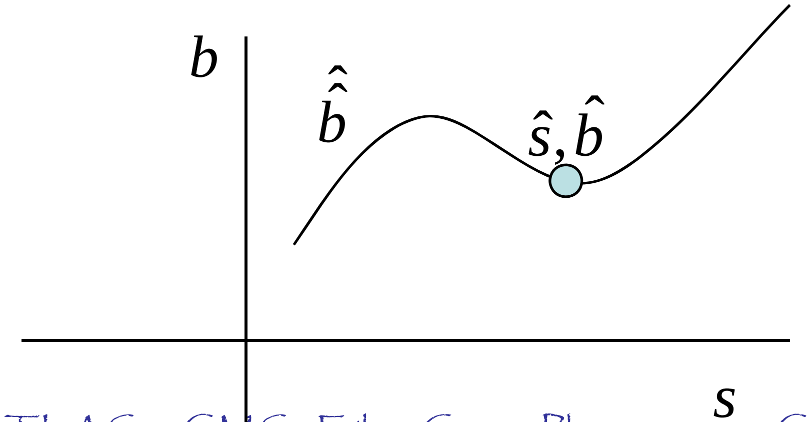
- Note that for every **s** this construction cover for $\hat{\hat{b}}$, in particular for the true value of **s**

The Profile Likelihood

- The recent introduction of the Profile Likelihood in HEP turned out to be a rediscovery of the MINOS Processor of MINUIT (**Fred James**) which was based on the Profile Likelihood…

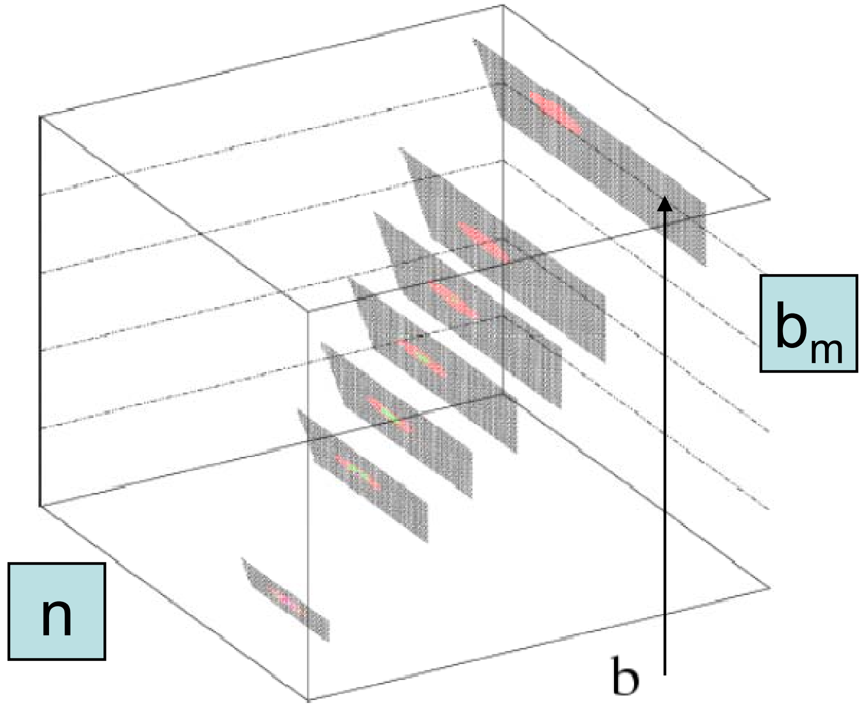
- The method allows the removal of the Nuisance parameters by replacing them with their conditional maximum likelihood estimates

$$Q(n,s)=\frac{L\left(n,b_m\mid s,\hat{\hat{b}}\right)}{L\left(n,b_m\mid \hat{s},\hat{b}\right)}$$



Full Projection Construction

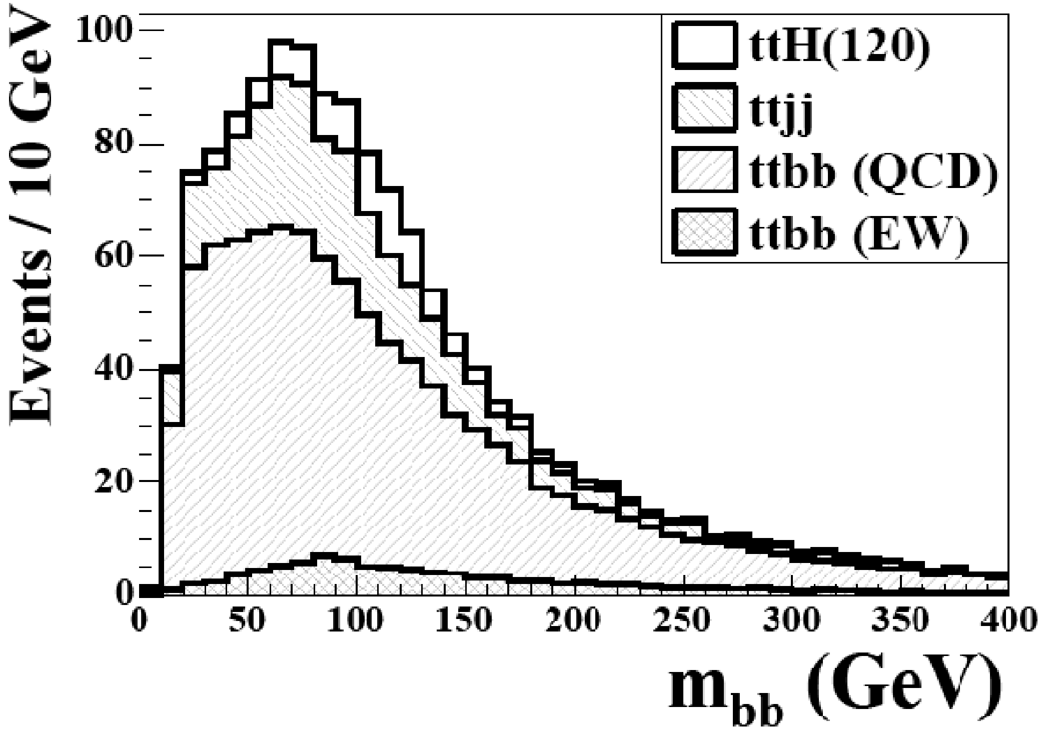
- FULL CONSTRUCTION
  - Complicated….
  - Cumbersome…
  - Computing Power Consumer
- Some say – not realistic…. The more dimensions, the less realistic….
- Go back to F&C construction… less dimensions… If you ensure a coverage for the most probable values of the Nuisance parameters and a full reconstruction for the parameters of interest… you should be OK….

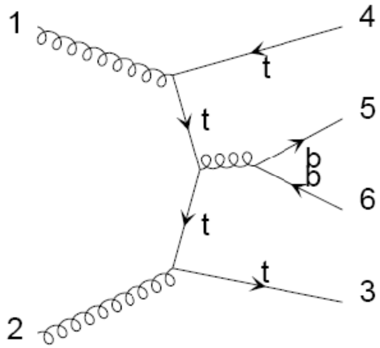


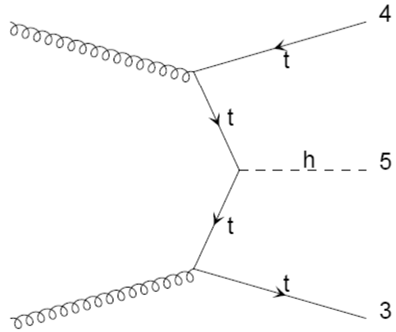
•See K. Cranmer Phystat 2003

Example: ttH→ttbb

- This is a very difficult analysis, here, with infinite luminosity one can never reach a discovery….







J. Cammin & M. Schumacher, ATL-PHYS-2003-024

Different Generators

- If two generators model different shapes… What should one do?
- So control samples are used… But that is not always a savior…..
- If one takes the difference in shapes as an uncertainty, say Δ, than the significance

$$s/\sqrt{b} \Rightarrow s/\sqrt{b(1+b\Delta^2)} \xrightarrow{L\to\infty} \frac{s/b}{\Delta}$$

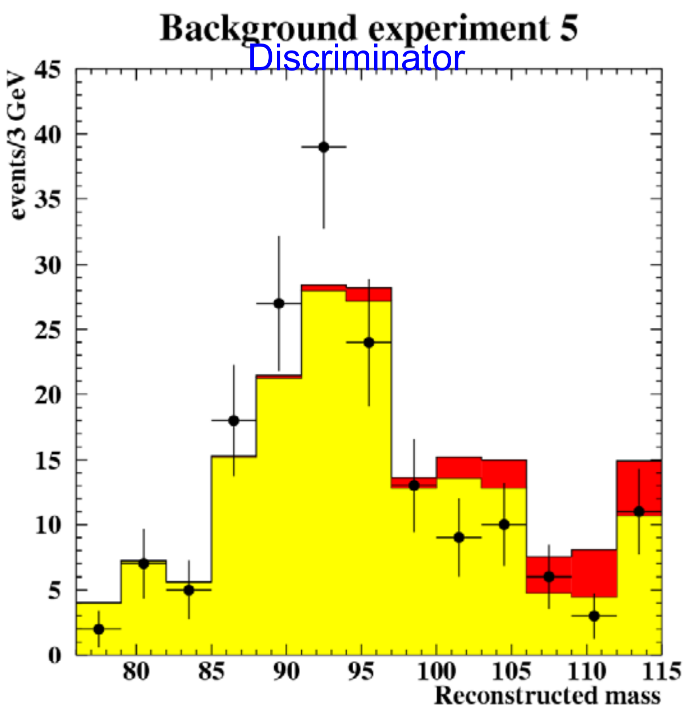
- $\frac{s/b}{\Delta} \geq 5 \to s/b \geq 0.5 \;\; for\, \Delta \sim 10\%$

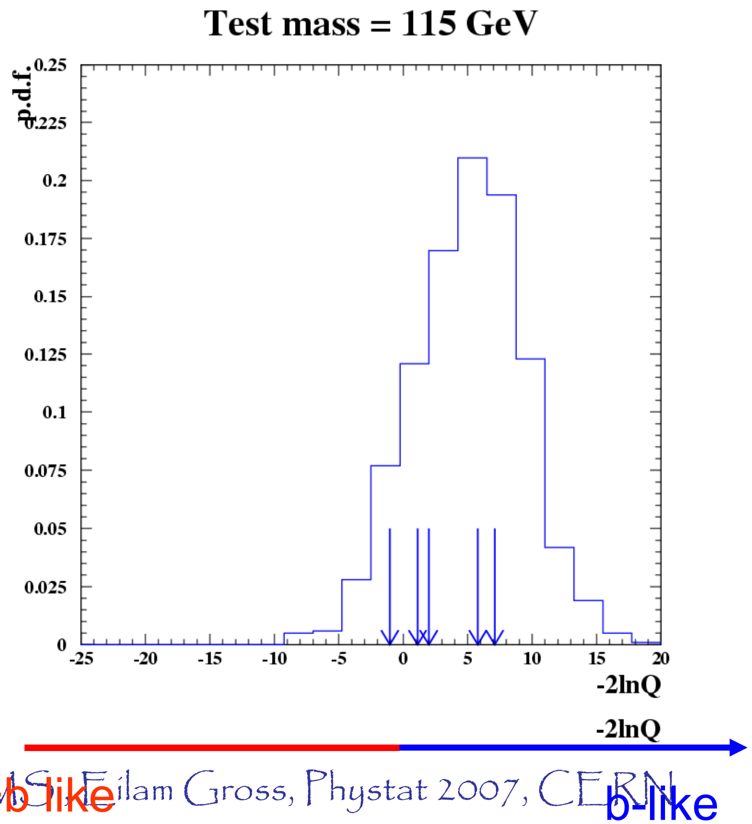
Example: Simulating BG Only Experiments

- The likelihood ratio, $-2\ln Q(m_H)$ tells us how much the outcome of an experiment is signal-like

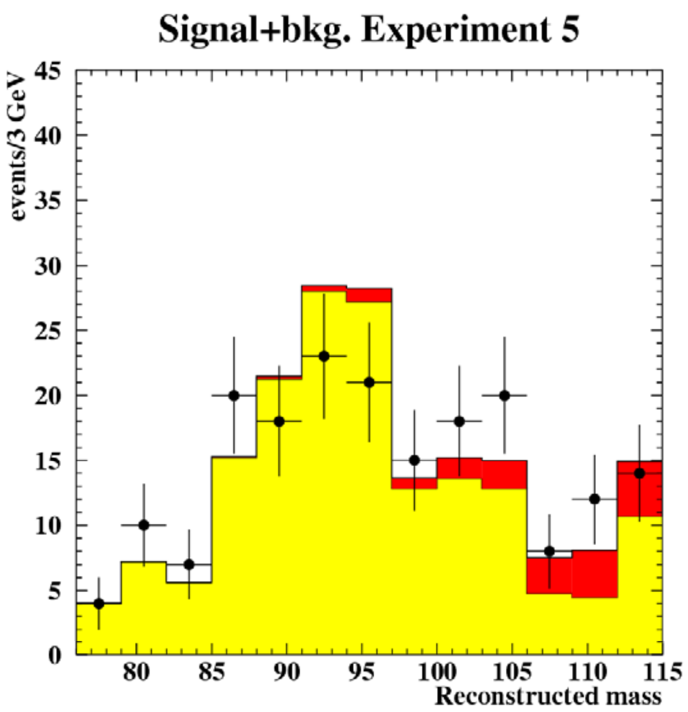
$$Q(m) = \frac{L(H_1)}{L(H_0)} = \frac{L\big(s(m)+b\big)}{L(b)}$$

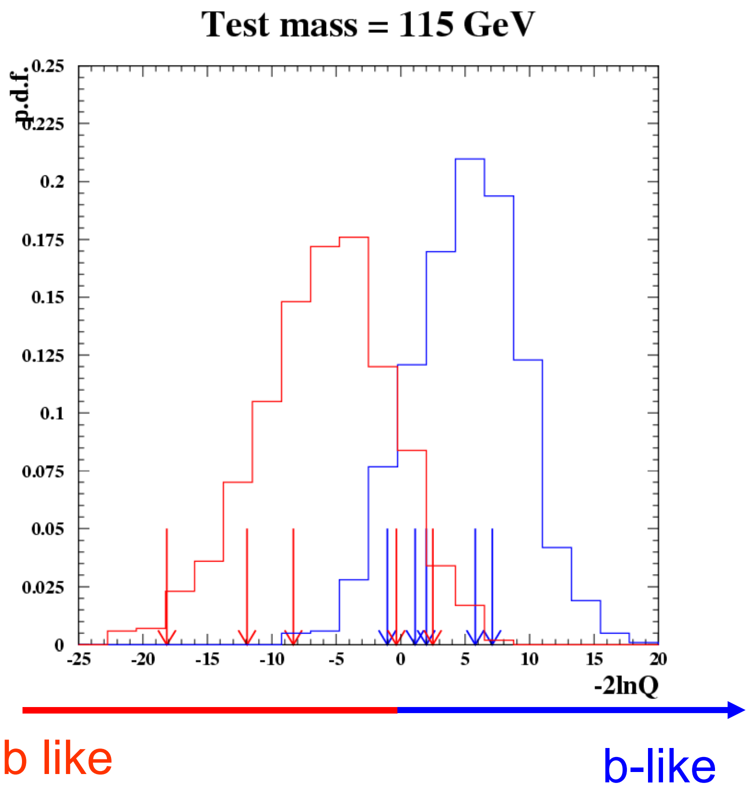
- NOTE, here the s+b pdf is plotted to the left!



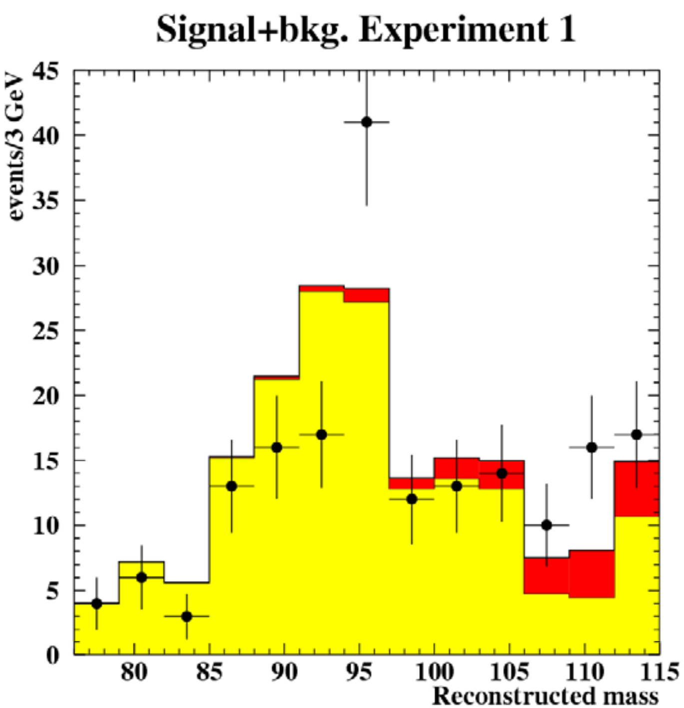


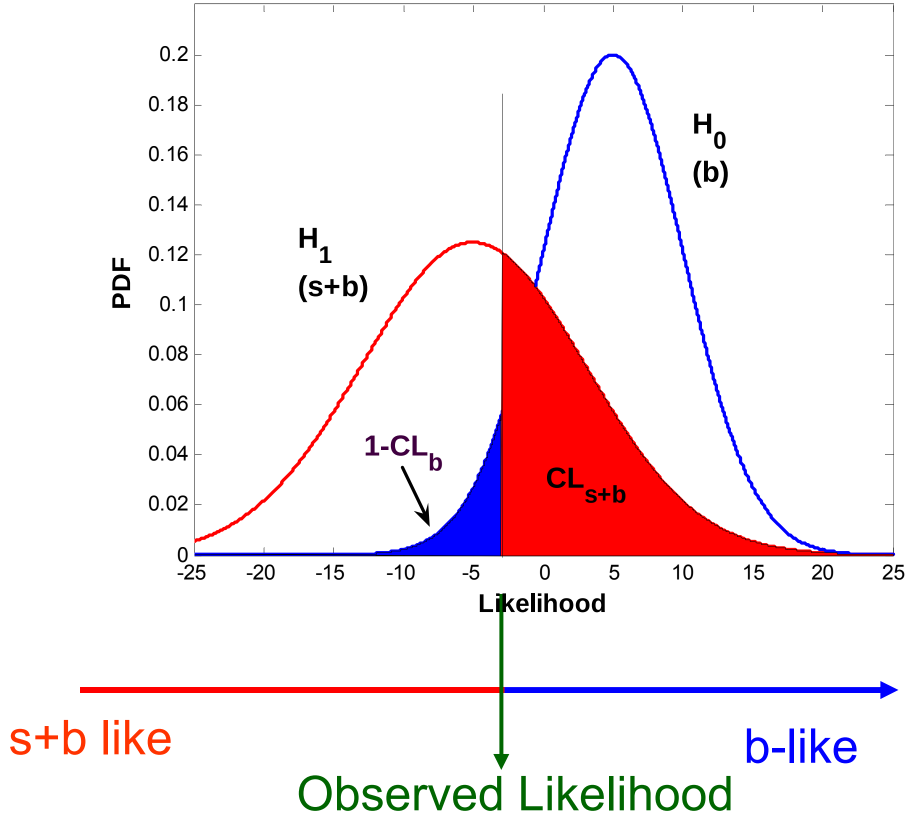
Example:
Simulating $S(m_H)$+b Experiments



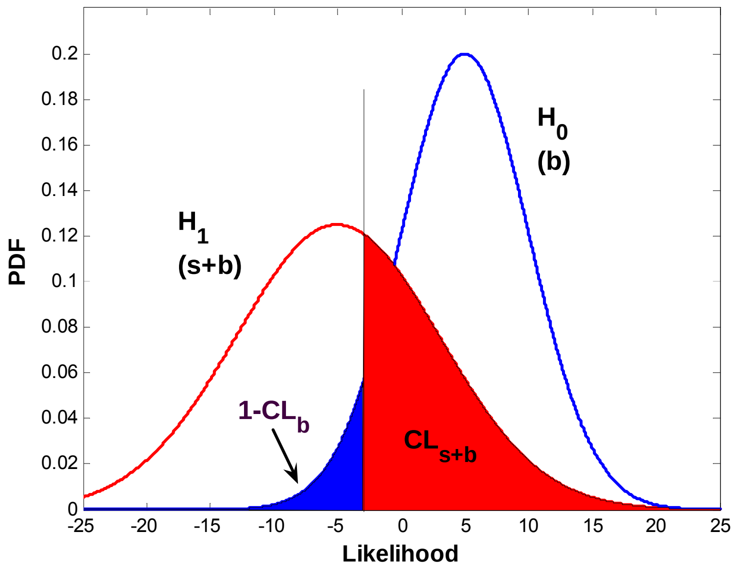


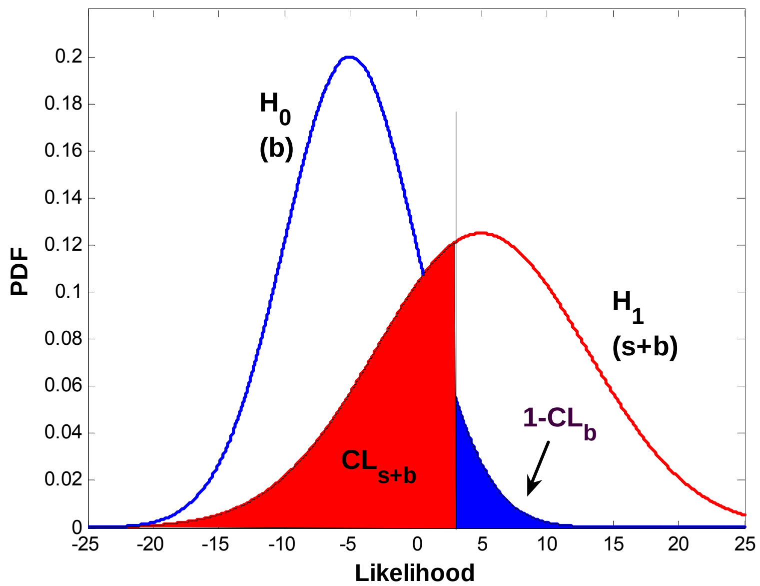
Example: Simulating $S(m_H)+b$ Experiments





Straighten Things Up





s+b like → b-like

b-like ← → s+b like

The CLs Method

- The CLs method used at LEP has in fact many ingredients a frequentist needs, except that it is criticized by the hardcore frequentists community…..
- It was developed to be used as a powerful tool to exclude the alternate hypothesis H_1 (a Higgs with a mass m) in favor of the (H_0) BG only null hypothesis.
- One constructs a test statistics Q based on the Neyman-Pearson lemma

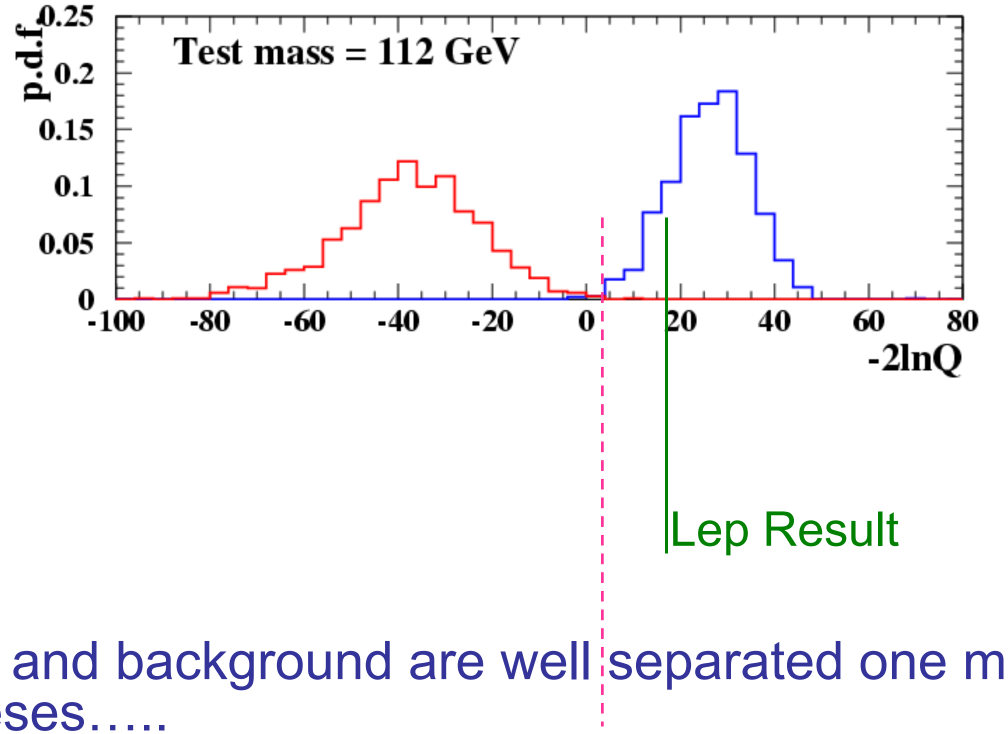
$$Q = \frac{P_{poisson}(data|s+b)}{P_{poisson}(data|b)} = \frac{L(s+b)}{L(b)}$$

- One calculates the **p-value** ($1-CL_b$) to assess the discovery sensitivity
- The rate of rejecting the Signal hypothesis using CL_s is lower then quoted (for small signals) therefore **the method undercovers FOR THE SIGNAL HYPOTHESIS.**
- How come?
  Being conservative is not always a virtue….

The Problem with the CL_s Method

- Suupose the limit using CLs+b is **m_H>116 GeV**
  CL_{s+b}<CL_s → the limit using CLs is **m_H>115 GeV (CONSERVATIVE)**
  - Suppose m_H=115.5 Gev
  - The CLs+b method rejected the 115.5 GeV Higgs in 5% of the experiments
  - The CLs method will fail to reject the 115.5 GeV often enough…. →UNDERCOVERAGE of the SIGNAL HYPOTHESIS
- The advocator will admit that the method will not exclude a small signal (to which the experiment is not sensitive) often enough
- Is it really that bad that a method undercover where Physics is sort of handicapped… (due to loss of sensitivity)?
- Alex Read when trying to apply the CL_s method for Neutrino oscillations measurements and discovery finds out the method cannot tell one possible signal point (in the Δm^2 vs. $\sin^2 2\theta$ plane) from another concluding that
  one should not push CLs beyond the level of exclusion and suggests to flip flop from CLs to F&C when the significance of a signal starts to show….
  (A. Read, Durham Phystat)

A Possible Anomaly



- If signal and background are well separated one might exclude both hypotheses…..
  i.e. makes a discovery and exclusion at the same time…..

The Flip Flop Way of an Experiment

- The most intuitive way to analyze the results of an experiment would be
- Construct a test statistics (for goodness of fit or for discriminating hypotheses) from the measure quantity x. Could be
  $Q(x) \sim L(x|H_0)$ if no clear signal hypothesis is known,
  or $Q(x) \sim L(x|H_1)/L(x_{obs}|H_0)$

- If the significance of the measured $Q(x_{obs})$, is less than 3 sigma, derive an upper limit (just looking at tables), if the result is >5 sigma (and some minimum number of events is observed….), derive a discovery central confidence interval for the measured parameter (cross section, mass….)
  …..

- This Flip Flopping policy leads to undercoverage: If we construct an a priori 90% Confidence acceptance interval [x1,x2] it turns out that there are true values of x such that $P(x_{true} \in [x1,x2]) < 90\%$

- Is that a problem?
  Some physicists say, for each experiment quote always two results, an upper limit, and a (central?) discovery confidence interval

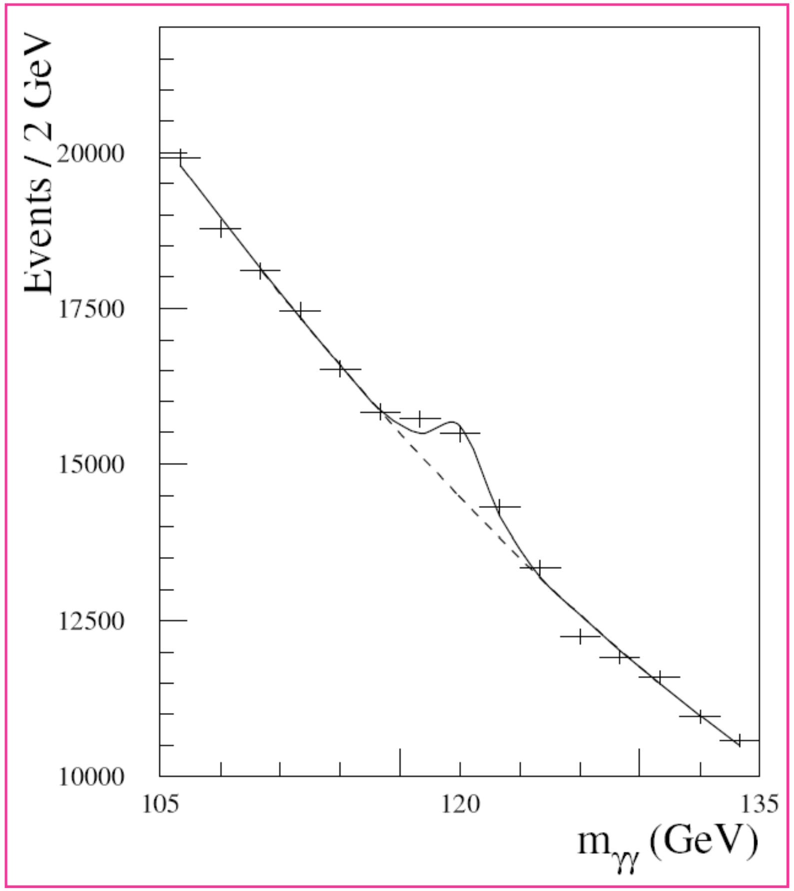
Conditional Confidence Intervals

- Physicists Habbit: If 5σ report a discovery and quote a 1σ confidence interval
- If less then 5σ reort a 95% upper limit
- An alternative to F&C unified approach (L. Demortier):
  - Define a 5σ control region a. Check the p-value… if $p<\alpha$, reort a discovery, if $p>\alpha$ calculate an upper limit.
  - To ensure coverage calculate the conditional Confidence Intervals (subject to your observation: Yes/No 5σ)

L. Demortier

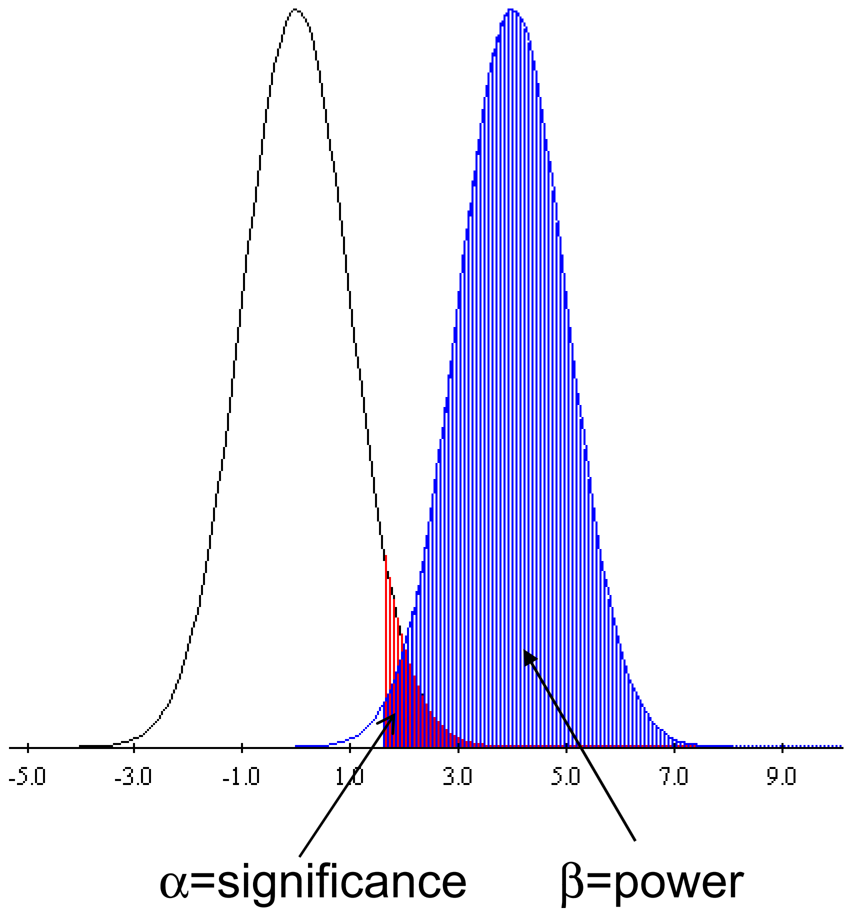
Looking for Guidelines on Subsequent Inference

- Example 2:
  - Search for a Gaussian resonance on top of a continuous smooth background
  - You make a decision according to the significance of the effect….. But when calculating the normalization (e.g.) you do not take into account the possibility that your decision was mistaken….



Which Analysis is Better

- To find out which of two methods is better plot the significance vs the power for each analysis method
- Given the significance, the one with the highest power is better



Look Elsewhere

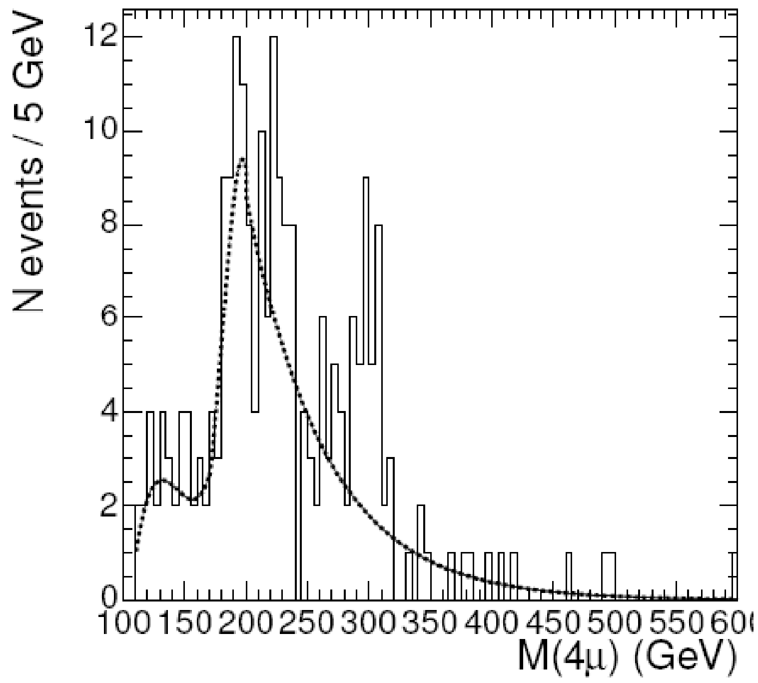
- To calculate a significance for a given value of m_H, we normally calculate the probability that a background fluctuation could have caused a greater or equal excess at that particular value of m_H (p-value).

- In a realistic search, one should consider the probability that a given fluctuation could occur anywhere within the mass range under study; This fact could have a nontrivial effect on the sensitivity of an analysis.

- We might claim that most spurious discoveries or observation were due to wrong inputs (Discovery of SUSY via mono-jets and the Altarelli cocktail…),

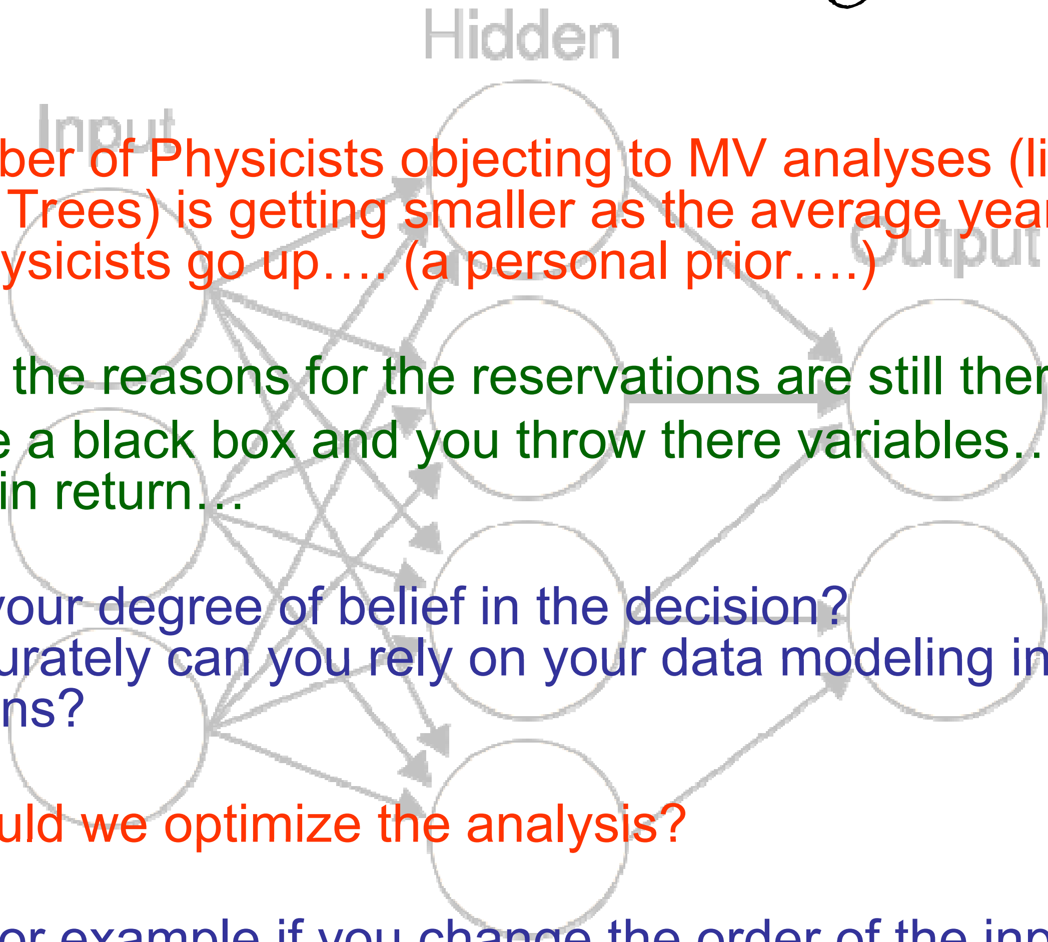
  Right example: A real fluctuation

See talk of Alexey Drozdetskiy



CMS TDR App. A

Multivariate Anlyses



- The number of Physicists objecting to MV analyses (like ANN, Decision Trees) is getting smaller as the average year of birth of the active physicists go up…. (a personal prior….)
- However the reasons for the reservations are still there….
- You have a black box and you throw there variables…. You get a decision in return…
- What is your degree of belief in the decision?
  How accurately can you rely on your data modeling in multi dimensions?
- How should we optimize the analysis?
- IN ANN for example if you change the order of the input variables, your decision might change….
  However, Physicists do not use the multilayer option… They do simple things