# Statistical Challenges of the LHC

**Kyle Cranmer**
Brookhaven National Laboratory

**September 13, 2005**
PhyStat2005, Oxford

Outline:

- The state of the art of frequentist hypothesis testing with no systematics
- Including systematics: translating confidence intervals to hypothesis testing
- Evaluation of the various methods
- Challenges of searches for beyond the standard model physics

The *Standard Model* of Particle Physics is a particular Quantum Field Theory that represents our best understanding of particles and their interactions.

The standard model is very predictive and has survived numerous precise tests over the years.

The only particle of the standard model that we have not observed is the Higgs boson.

Despite its success, we have reason to believe that it is not the whole story: we expect that there will be some deviation from the standard model near the TeV energy scale.

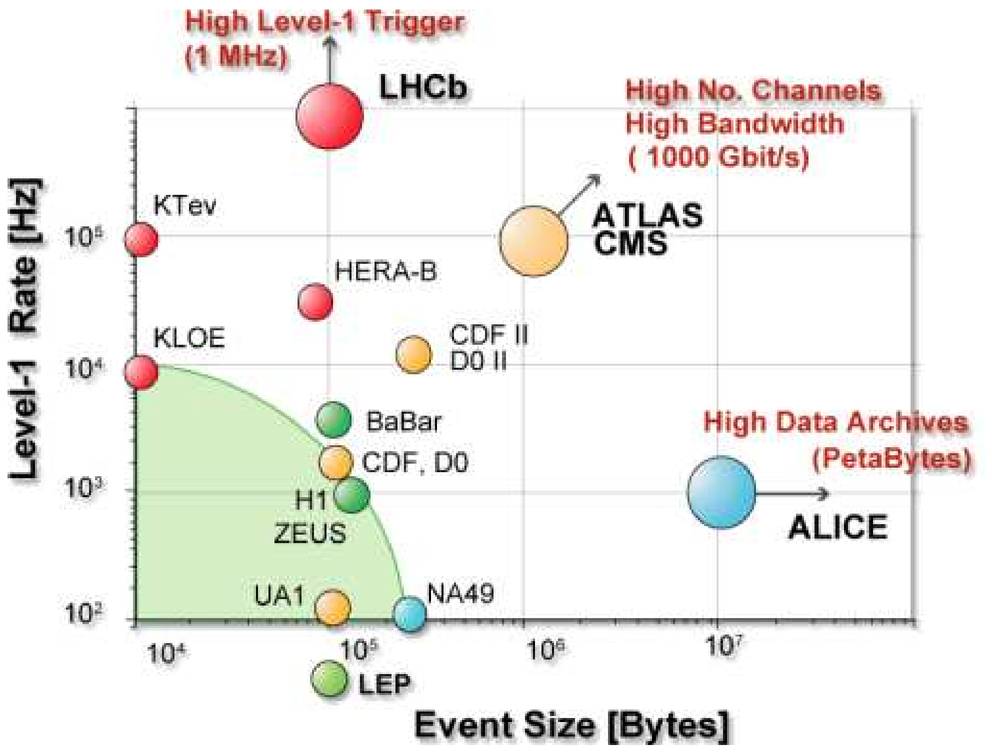
Physics beyond the standard model includes: SuperSymmetry (SUSY), extra space-time dimensions, new high-mass resonances, etc.

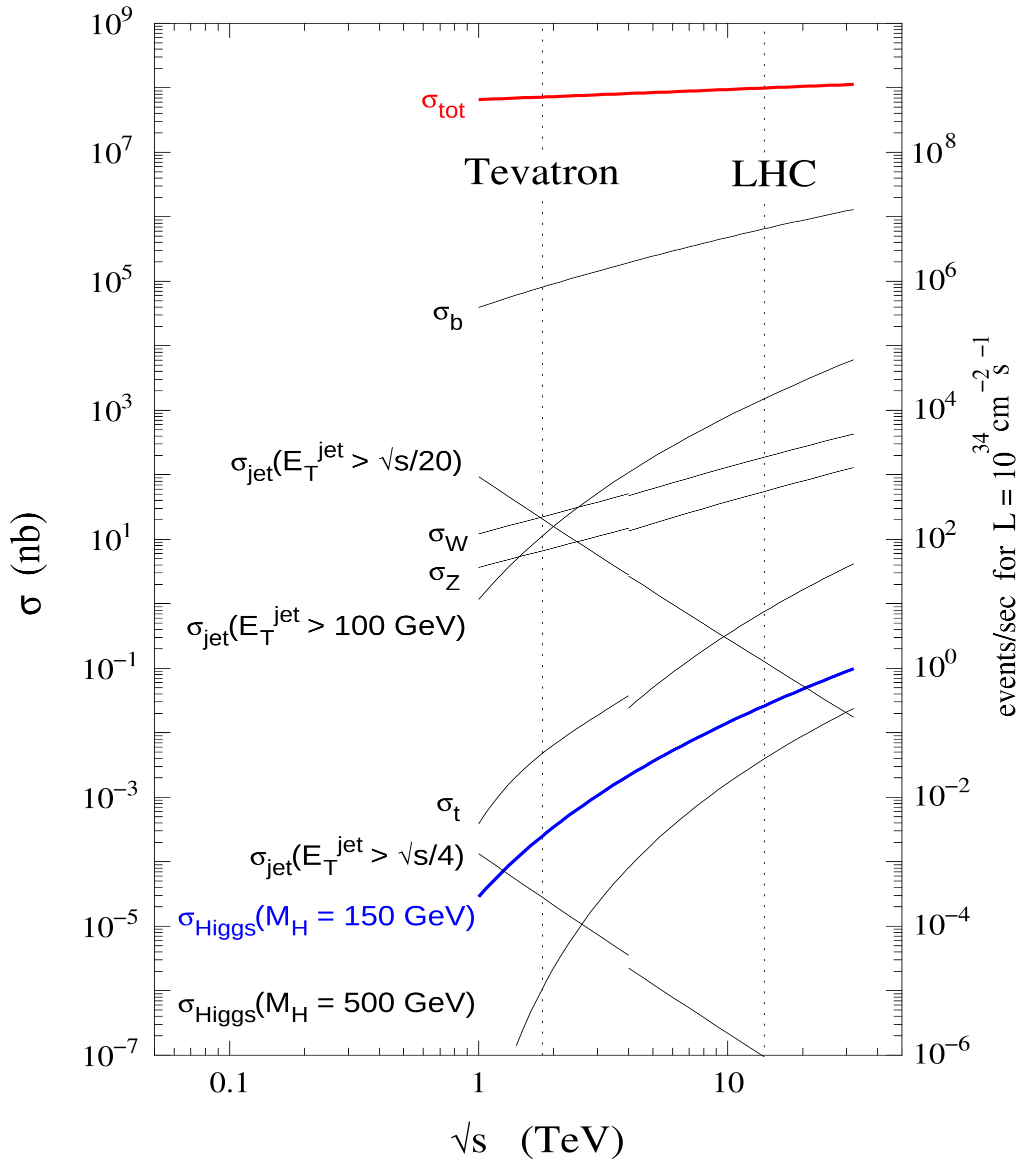
The Large Hadron Collider (LHC) at CERN and the two large multi-purpose detectors (ATLAS and CMS) have been built specifically to find the standard model Higgs boson (if it exists) and explore the theoretical landscape of beyond the standard model.

At the LHC, bunches of protons will collide 40 million times per second!

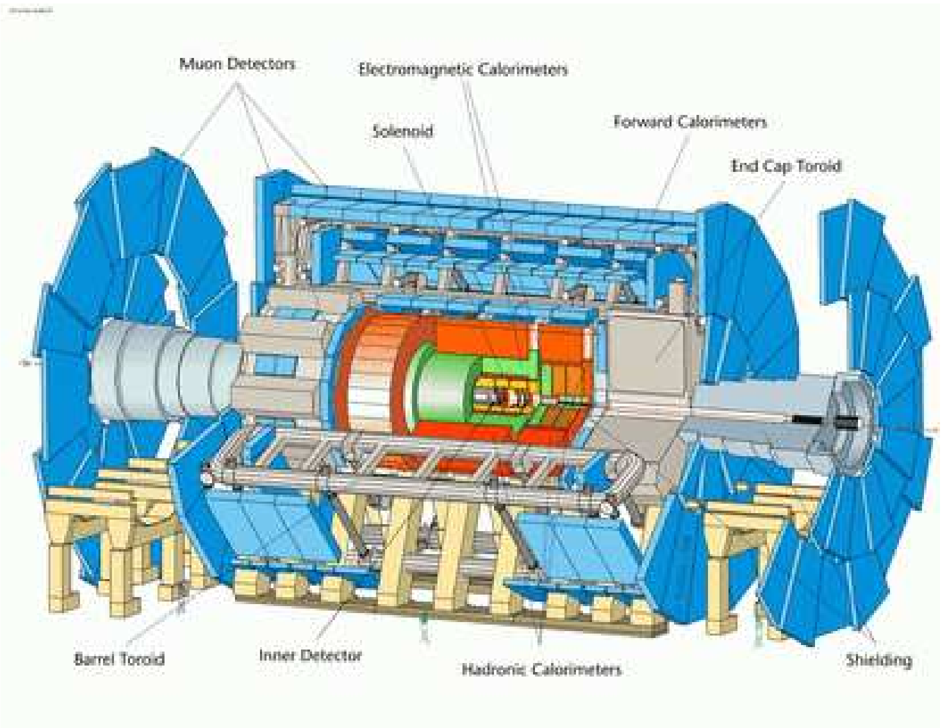
Most interactions are not interesting.

We store about 200 events/s, each about 2MB

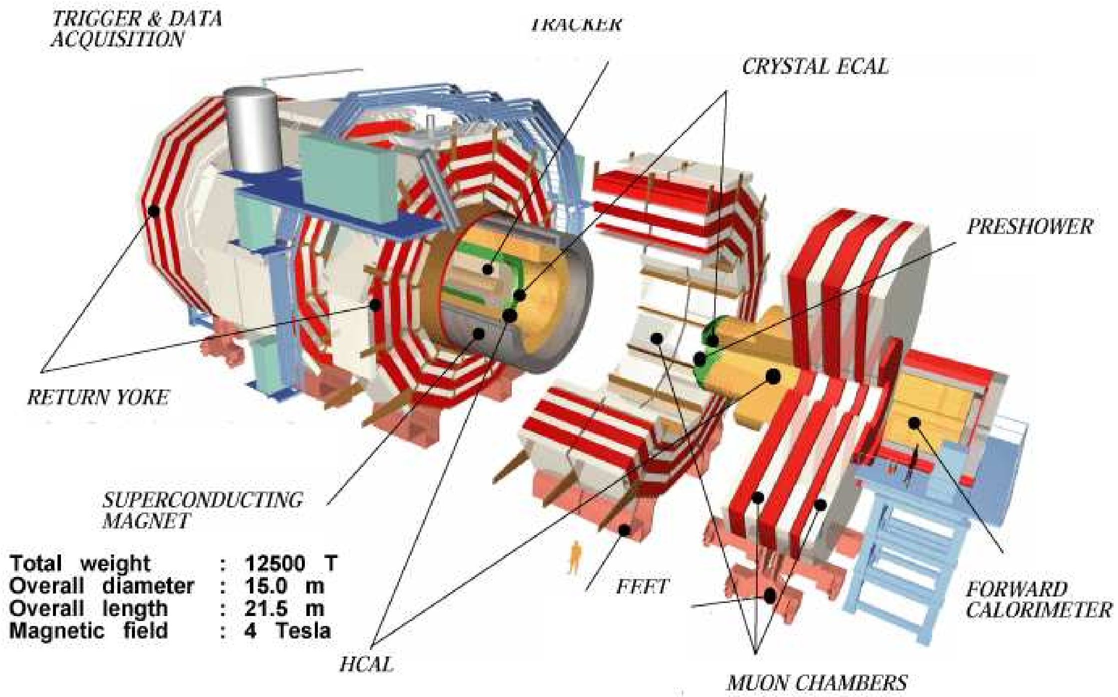


## The ATLAS & CMS Detectors




Both experiments have ∼2000 collaborators!




**ATLAS**

- Length ≈ 40 m
- Radius ≈ 10 m
- Weight ≈ 7000 tons
- Electronic Channels ≈ $10^8$

The ATLAS & CMS are a multipurpose detectors...

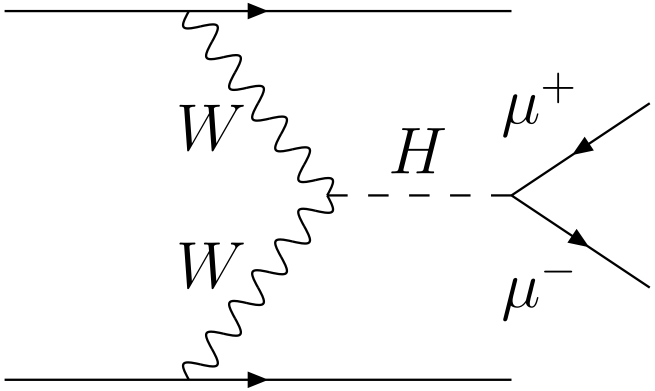
flexible enough for the surprises which may lie ahead!


C

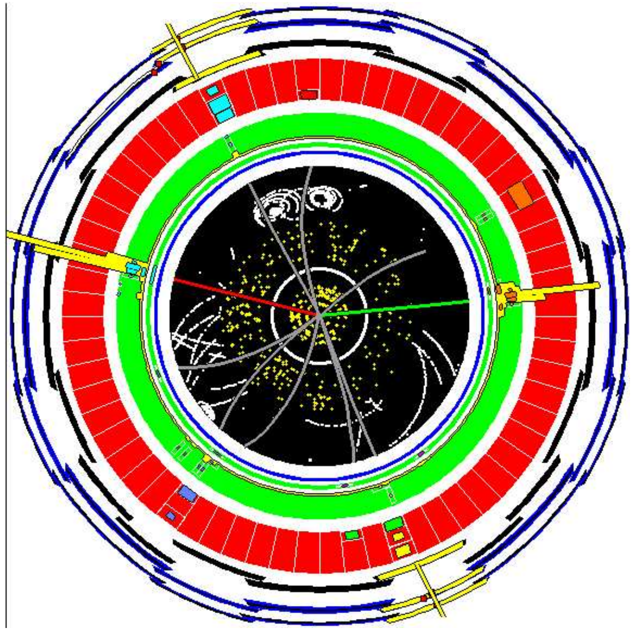
## *Prediction via Monte Carlo Simulation*

The enormous detectors are still being constructed, but we have detailed simulations of the detectors response.

$$L(x|H_0) =$$




$$\oplus$$



The advancements in theoretical predictions, detector simulation, tracking, calorimetry, triggering, and computing set the bar high for equivalent advances in our statistical treatment of the data.

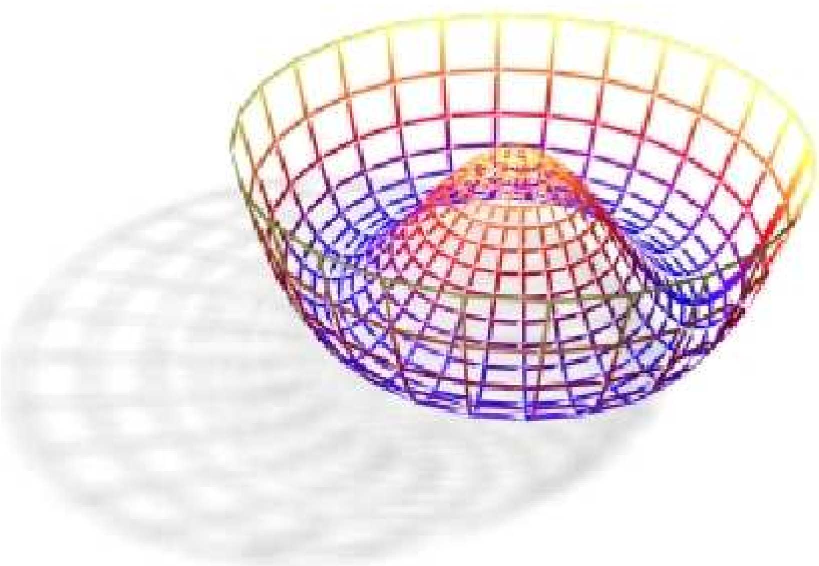
The theoretical model for the standard model **Higgs** only has **one free parameter**:

- The mass of the Higgs boson $m_H$

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$$

With $m_H$ specified, the theory predicts:

- production rates
- angular distributions
- branching ratios $H \rightarrow ZZ, WW, \gamma\gamma, \tau\tau$

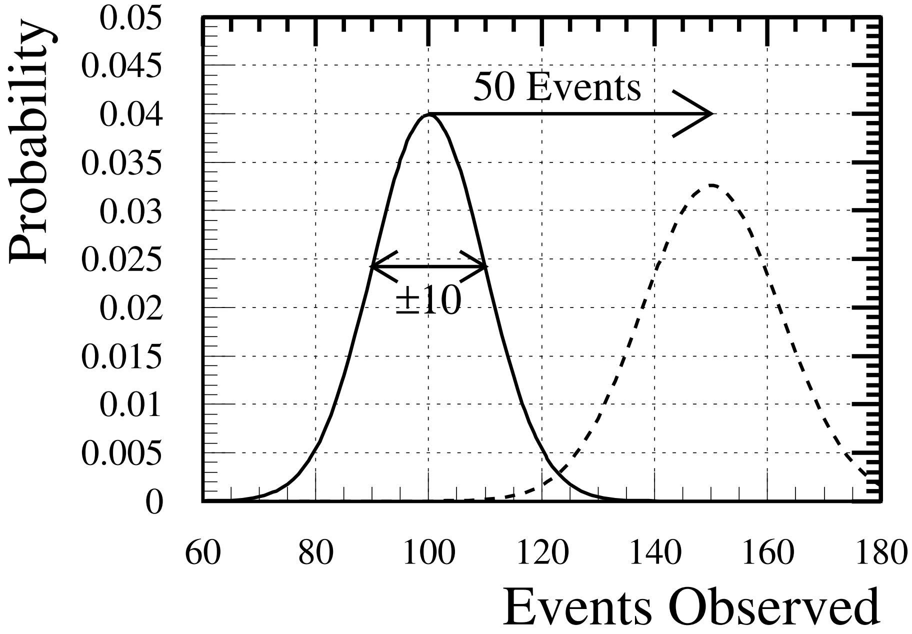


**Supersymmetry** is more of a framework than a theory *per se*
(we don't know how the symmetry is broken)

MSSM is a parametrized theory of all soft-SUSY breaking terms: **has 105 parameters!**

Other models like mSUGRA, mGMSB, mAMSB have $\sim$ 4 parameters

SUSY searches focused more on inclusive signatures for discovery and mass measurements for parameter determination




Only works if we know background perfectly

In a "number counting" experiment one considers all events as equally "signal like"

A more powerful method would take into account event characteristics

The $5\sigma$ discovery threshold means that we claim discovery when the Higgs is not there (commit a Type I Error) with a probability of $2.85 \cdot 10^{-7}$.

For $b = 100$ events we need $x \geq 154$ to claim a discovery

**Sketch of a number counting search:**

1. Define signal-like region in data ("cuts")
2. Predict $b$ events from background process
3. Predict $s$ events from signal process
4. Observe $x$ events
5. Define $p = Pois(n \geq x|b)$.
6. Require $p < \alpha$ to claim a discovery

1992 - R. D. Cousins and V. L. Highland, Nucl. Instr. and Meth. A320, 331 (1992).
Including background uncertainty with Bayesian smearing

1997 - G. J. Feldman and R. D. Cousins, Phys. Rev. D57, 3873 (1998).
The Neyman construction with the Likelihood ratio as an ordering rule.

1997 - A. Read's DELPHI note on the Likelihood Ratio, extending Poisson case to use discriminating variables

1999 - T. Junk hep-ex/9902006, Generalized the "Cousins Highland" method for including background systematic uncertainty

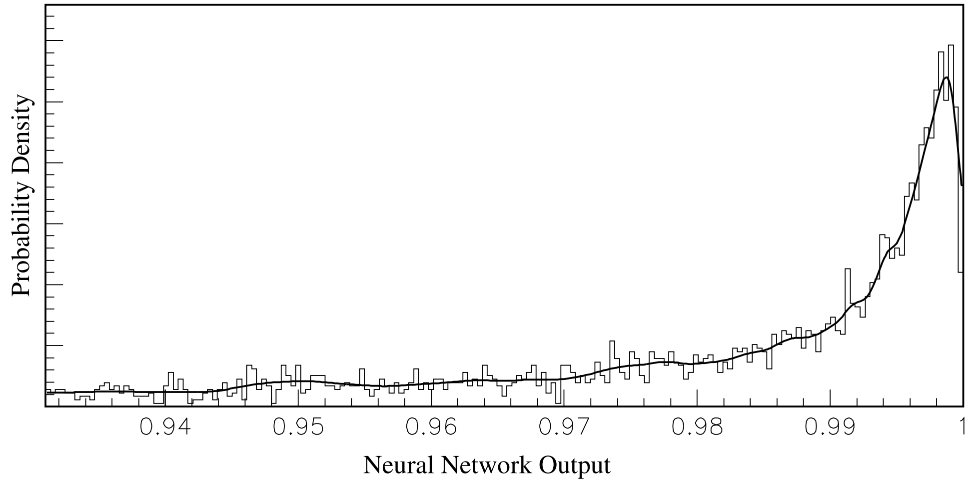
2000 - The first workshop on confidence limits was held at CERN

2000 - H. Hu & J. Nielsen [physics/9906010] Analytic confidence level calculations using likelihood ratio and Fourier Transform

2001 - K. Cranmer, Comput. Phys. Commun. 136, 198 (2001)
Using kernel estimation to describe p.d.f's of arbitrary distributions for use in likelihood ratio

I was only peripherally involved in the LEP Higgs effort, but I did make one contribution, KEYS, which was used to define p.d.f of discriminating variables

Most channels used at least one "discriminating variable" to improve the discrimination against background




Often these variables were the output of Neural Networks or some other multivariate algorithm

It's hard to parametrize these shapes, so used Kernel Estimation instead.

$$\hat{f}_1(x) = \sum_i^n \frac{1}{nh(x_i)} K\left(\frac{x - x_i}{h(x_i)}\right)$$

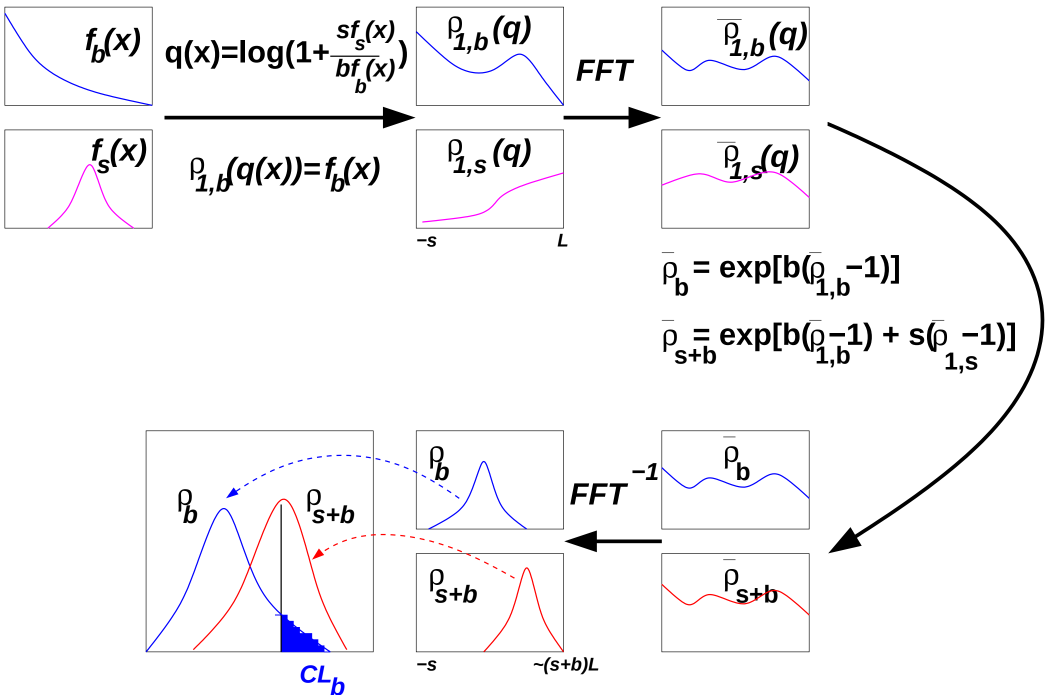
$$h(x_i) = \left(\frac{4}{3}\right)^{1/5} \sqrt{\frac{\sigma}{\hat{f}_0(x_i)}} n^{-1/5}$$

KEYS was used by LEP Higgs working group and BaBar

LEP Higgs Working group developed formalism to combine channels and take advantage of discriminating variables in the likelihood ratio.

$$Q = \frac{L(x|H_1)}{L(x|H_0)} = \frac{\prod_i^{N_{chan}} Pois(n_i|s_i+b_i) \prod_j^{n_i} \frac{s_i f_s(x_{ij}) + b_i f_b(x_{ij})}{s_i+b_i}}{\prod_i^{N_{chan}} Pois(n_i|b_i) \prod_j^{n_i} f_b(x_{ij})}$$

$$q = \ln Q = -s_{tot} \sum_i^{N_{chan}} \sum_j^{n_i} \ln\left(1 + \frac{s_i f_s(x_{ij})}{b_i f_b(x_{ij})}\right)$$




Hu and Nielsen's CLFFT used Fourier Transform and exponentiation trick to transform the log-likelihood ratio distribution for one event to the distribution for an experiment

Cousins-Highland was used for systematic error on background rate.

Getting this to work at the LHC is tricky numerically because we have channels with $n_i$ from 10-10000 events (physics/0312050)

**Nothing is Poisson**, we weight events with discriminating variable pdfs

Not all shapes can be parametrized, we use kernel estimation for **non-parametric estimates**

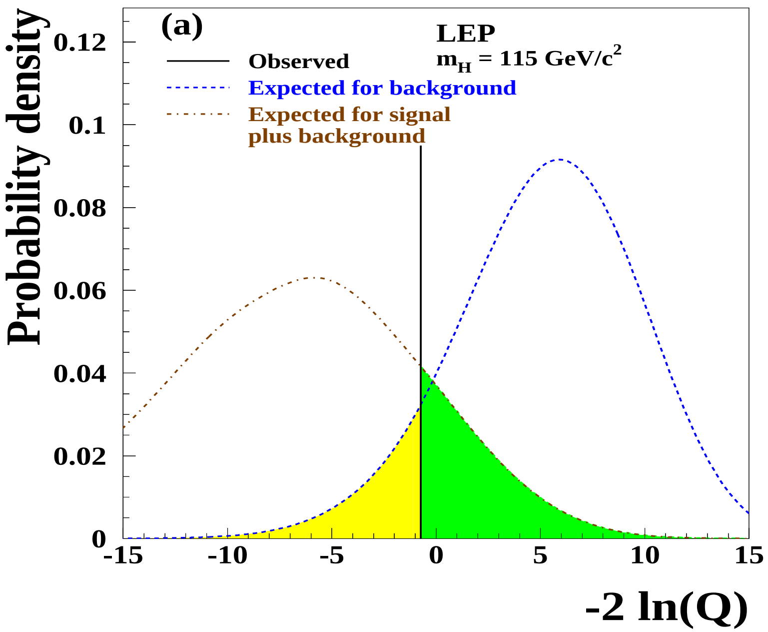
We use **numerical techniques** to evaluate $p$-value

Combining results requires the use of compatible (if not identical) methods for a wide range of situations

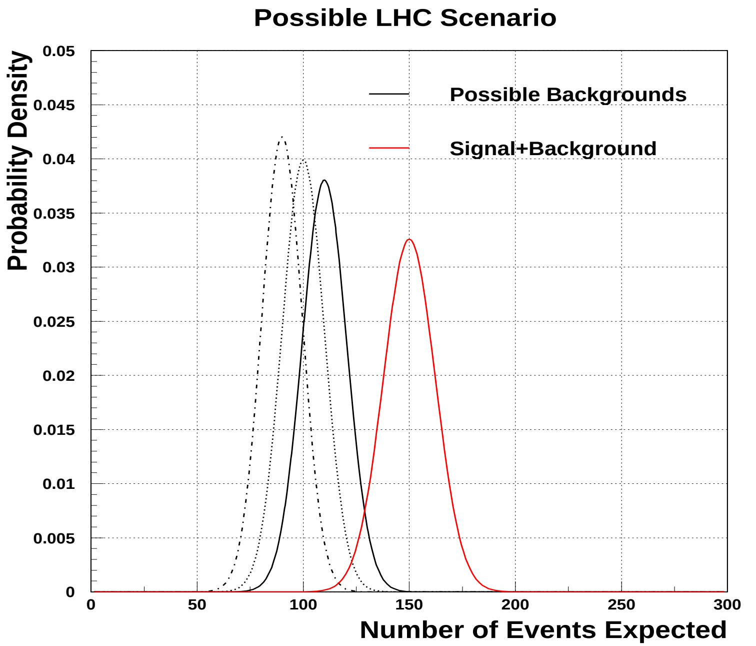
To make numerical procedures work for a wide range of problems, we need to know **asymptotic scaling behavior**

At LEP (and to some extent Tevatron)
- small sample sizes,
- small systematic errors,
- and emphasis was on **confidence limits**

At the LHC
- large range of samples sizes,
- large systematic error,
- and emphasis will be on **discovery**




Overlapping distributions
Width dominated by limited # of events




Well-separated distributions
Width dominated by background uncertainty

My focus in recent years has been on:

- migrating the LEP Higgs statistical techniques to the LHC (mainly for ATLAS)
- translating the work on frequentist confidence intervals with systematic errors (nuisance parameters) into the language of hypothesis testing
- discovering the situations where the common techniques go most astray
- incorporating these issues into the ATLAS analysis model

My point in the remaining slides is that

- we basically understand the issues and the formalism, but
- most of the existing tools are not sufficient for LHC discoveries

## How Confidence Intervals Differ from Hypothesis Testing

Confidence Intervals from the Neyman Construction are "inverted" Hypothesis Tests –
*I don't deny they are formally equivalent*.

However, there are practical differences between 90% intervals and $5\sigma$ discoveries

1. There is not always a continuous family of (physically motivated) models $L(x|\mu)$ that ties together $H_0 = \mu_0$ and $H_1 = \mu_1$
   - $H_0 =$ spin 1 $Z'$ vs. $H_1 =$ spin2 $Z'$
   - $H_0 =$ S.M. no Higgs vs. $H_1$ S.M. with Higgs
   - physically motivated may just mean "we know how to generate the Monte Carlo"
2. Searches don't really care about small over-coverage from discreteness of small signals
3. One point of the interval $H_0$ is particularly important, we really care about coverage of this point. (Important for discussion on projecting-out nuisance parameters)
4. Much harder to trust an approximation out to $5\sigma$, need numerical evaluation

Consider the case of a

- Poisson process $X$ in the signal like region
- A sideband measurement $Y$ used to predict the background
- And a background extrapolation from the sideband via $Y/\tau$

The likelihood for a given signal rate $\mu$ and background rate $b$ is thus,

$$L(X, Y|\mu, b) = Pois(X|\mu + b) \cdot Pois(Y|\tau b)$$

or

$$L(X, Y|\mu, b) = Pois(X|\mu + b) \cdot G(Y|\tau b, \sqrt{\tau b})$$

Let's consider the case $b_{true} = 100$ & $\tau = 1$ and ask what value of $x_{crit}(y)$ is necessary to reject the null hypothesis $\mu = 0$ when $y$ is measured

At PhyStat2003, J. Linnemann covered several common approaches to this problem in HEP and Astrophysics. Instead of finding $x_{crit}$ he asked "how significant is a given $x$"

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Non = x | 4 | 6 | 9 | 17 | 50 | 67 | 200 | 523 | 167589 | 498426 | 2119449 | |
| Noff = y | 5 | 18.78 | 17.83 | 40.11 | 55 | 15 | 10 | 2327 | 1864910 | 493434 | 23671193 | |
| $\alpha$ | 0.2 | 0.0692 | 0.2132 | 0.0947 | 0.5 | 2.0 | 10.0 | 0.167 | 0.0891 | 1.000 | 0.0891 | |
| $b = \alpha y$ | 1.0 | 1.3 | 3.8 | 3.8 | 27.5 | 30.0 | 100.0 | 388.6 | 166213 | 493434 | 2109732 | |
| s = Non - b | 3.0 | 4.7 | 5.2 | 13.2 | 22.5 | 37 | 100 | 134.4 | 1376 | 4992 | 9717 | |
| $\delta b$ | 0.45 | 0.3 | 0.9 | 0.6 | 3.71 | 7.75 | 31.6 | 8.1 | 121.7 | 702.4 | 433.6 | |
| $\delta b$/b | 0.447 | 0.231 | 0.237 | 0.158 | 0.135 | 0.258 | 0.316 | 0.0207 | 0.000732 | 0.00142 | 0.000206 | |
| Reported p | | .0030 | .027 | 2.0E-06 | | | | | | | | |
| Reported Z | | 2.7 | 1.9 | 4.6 | 3.0 | 3.0 | | 5.9 | 3.2 | 5.0 | 6.4 | |
| Recommended: | | | | | | | | | | | | |
| $Z_{Bi}$ Binomial | **1.66** | **2.63** | **1.82** | **4.46** | **2.93** | **2.89** | **2.20** | **5.93** | **3.23** | **5.01** | **6.40** | 0 |
| $Z_\Gamma$ Bayes Gamma | **1.66** | **2.63** | **1.82** | **4.46** | **2.93** | **2.89** | **2.20** | **5.93** | * | * | * | 0 |
| Reasonable: | | | | | | | | | | | | |
| $Z_N$ Bayes Gauss (HEP) | 1.88 | 2.71 | 1.94 | 4.55 | 3.08 | *3.44* | *2.90* | **5.93** | **3.23** | **5.02** | **6.40** | .28 |
| $Z_0$ $\surd + 3/8$ | 1.93 | 2.66 | 1.98 | 4.22 | 3.00 | 3.07 | 2.39 | 5.86 | **3.23** | **5.01** | **6.40** | .15 |
| $Z_L$ L Ratio (GRA) | 1.95 | 2.81 | 1.99 | 4.57 | 3.02 | 3.04 | 2.38 | **5.93** | **3.23** | **5.01** | **6.41** | .14 |
| Not Recommended: | | | | | | | | | | | | |
| $Z_9 = s/\sqrt{\alpha(N_{on} + N_{off})}$ | *2.24* | *3.59* | 2.17 | *5.67* | 3.11 | **2.89** | **2.18** | 6.16 | **3.23** | **5.01** | **6.41** | .52 |
| $Z_5 = s/\sqrt{N_{on} + \alpha^2 N_{off}}$ | 1.46 | *1.90* | 1.66 | *3.17* | 2.82 | 3.28 | *2.89* | *5.54* | **3.22** | **5.01** | **6.40** | .93 |
| $Z_{5'} = s/\sqrt{\alpha(1+\alpha)N_{off}}$ | *2.74* | *3.99* | *2.42* | *6.47* | *3.50* | *3.90* | *3.02* | 6.31 | **3.23** | **5.03** | **6.41** | .53 |
| Ignore $\delta b$: | | | | | | | | | | | | |
| $Z_P$ Poisson: ignore $\delta b$ | 2.08 | 2.84 | 2.14 | 4.87 | *3.80* | *5.76* | *7.72* | *6.44* | 3.37 | *7.09* | 6.69 | 1.9 |
| $Z_3$ Fraser-Reid $\approx Z_P$ | 2.07 | 2.84 | 2.14 | 4.87 | *3.80* | *5.76* | *(8.95)* | *6.44* | 3.37 | 6.09 | 6.69 | 2.2 |
| $Z_{sb} = s/\sqrt{b}$ | *3.00* | *4.12* | *2.67* | *6.77* | *4.29* | *6.76* | *10.00* | *6.82* | 3.38 | *7.11* | 6.69 | 2.9 |
| Unsuccessful Hacks: | | | | | | | | | | | | |
| Poisson: Nb $\rightarrow$ b + $\delta b$ | 1.56 | 2.46 | 1.64 | **4.47** | 3.04 | *4.24* | *5.51* | 6.01 | 3.07 | *6.09* | **6.39** | 1.1 |
| s / $\sqrt{b + \delta b}$ | *2.49* | *3.72* | *2.40* | *6.29* | *4.03* | *6.02* | *8.72* | *6.75* | 3.37 | *7.10* | 6.69 | 2.4 |

The recommended Binomial method can't be directly applied to non-Poisson case

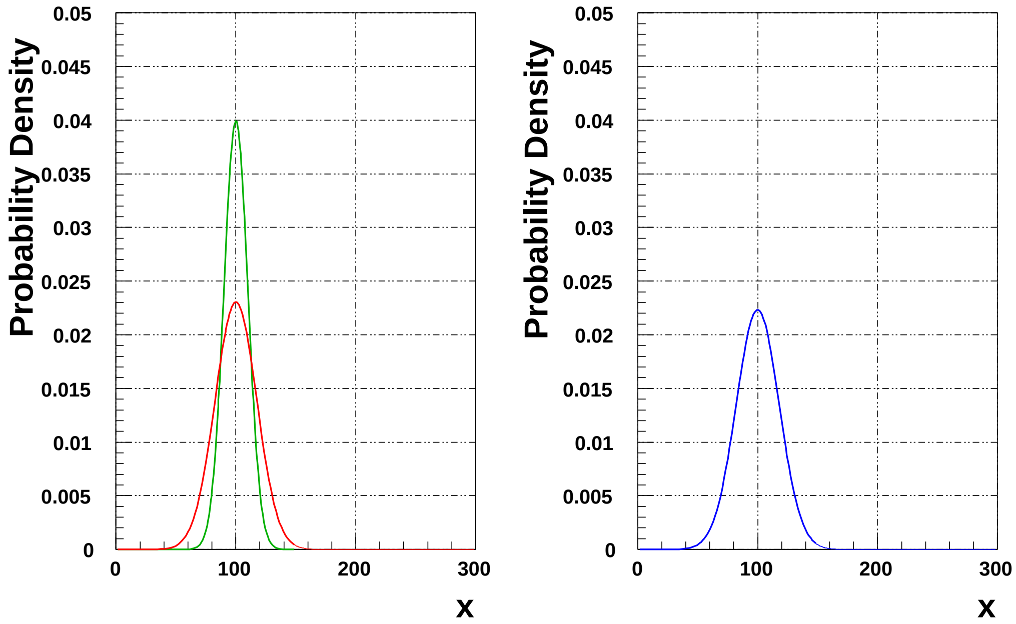
The Bayesian $Z_N$ (Cousins-Highland) and $Z_\Gamma$ (probably) undercover significantly (next)

Let's consider the generalization of the Binomial method and likelihood ratio...

LEP-Higgs used the Cousins-Highland method for systematic uncertainty on background rate

The Cousins-Highland method integrates-out $b$

$$L(x|H_0, Y) = \int_b db L(x|b) L(b|Y)$$




But it uses a Bayesian notion $L(b)$

$$L(b|Y) = \frac{L(Y|b)\ L(b)}{L(Y)}$$

The Tevatron also tends to use Bayesian techniques for background uncertainty

Method does well for small background uncertainty, but ...

**In my test with $L(Y|b) = G(\mu = 100, \sigma = 10)$ the $5\sigma$ critical region defined by the Cousins-Highland method had a rate of Type I error equivalent to $4.14\sigma$: significant under-coverage. Need $x_{crit} = x_{crit}^{CH} + 17$**

In the last few years, many people have looked at this section from Kendall's chapter on Likelihood ratio tests & test efficiency

| Variable | Meaning |
|---|---|
| $\theta_r$ | physics parameters |
| $\theta_s$ | nuisance parameters |
| $\hat{\theta}_r, \hat{\theta}_s$ | unconditionally maximize $L(x\|\hat{\theta}_r, \hat{\theta}_s)$ |
| $\hat{\hat{\theta}}_s$ | conditionally maximize $L(x\|\theta_{r0}, \hat{\hat{\theta}}_s)$ |

$$(H_0 : \theta_r = \theta_{r0})$$
$$(H_1 : \theta_r \neq \theta_{r0})$$

*Now consider the Likelihood Ratio*

$$l = \frac{L(x|\theta_{r0}, \hat{\hat{\theta}}_s)}{L(x|\hat{\theta}_r, \hat{\theta}_s)}$$

*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.*

There are at least three different methods that use this as a starting point

# Feldman-Cousins with Nuisance Parameters: "Profile Construction"

## Nuisance Parameters

- A nuisance parameter is an unknown parameter whose value is not of interest, but for which coverage must be provided for all possible values.
- In this talk I will be mainly concerned with the true rate of background production as a nuisance parameter.
- Obtaining exact coverage for nuisance parameters is a cumbersome procedure at best, and computationally impossible in complicated cases. Therefore, statisticians often use the approximate procedure suggested by Kendall and Stuart of eliminating the nuisance parameters by maximizing the likelihood with respect to them.

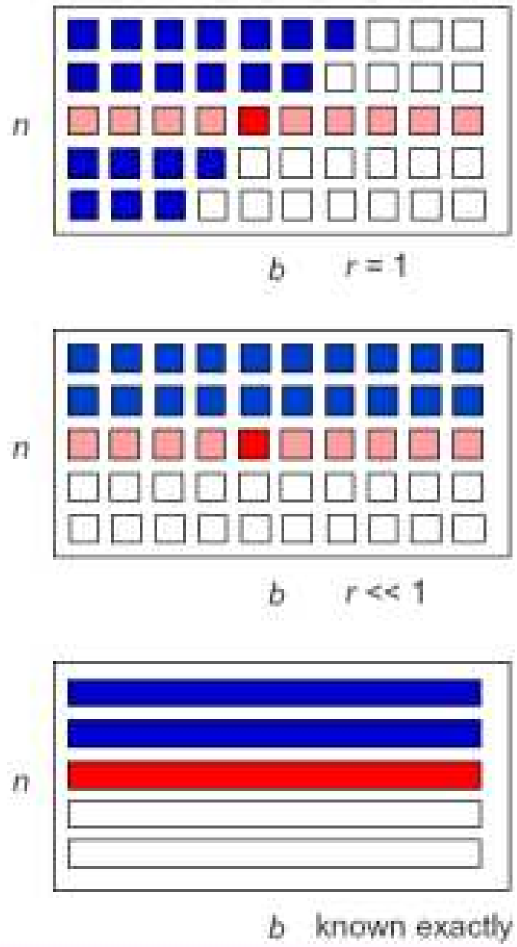
$$l(x,\theta_{r_0}) = \frac{L(x\,|\,\theta_{r_0},\hat{\hat{\theta}}_s)}{L(x\,|\,\hat{\theta}_r,\hat{\theta}_s)}.$$

The idea is that if one covers for $\hat{\hat{\theta}}_s$, the values most favorable to $\theta_{r_0}$, then one is likely to cover for all $\theta_s$. Our preliminary studies show that this is true to a high degree.

- The maximizations can be done analytically in simple cases, and numerically in more complex cases.

Gary Feldman 10 Fermilab Workshop

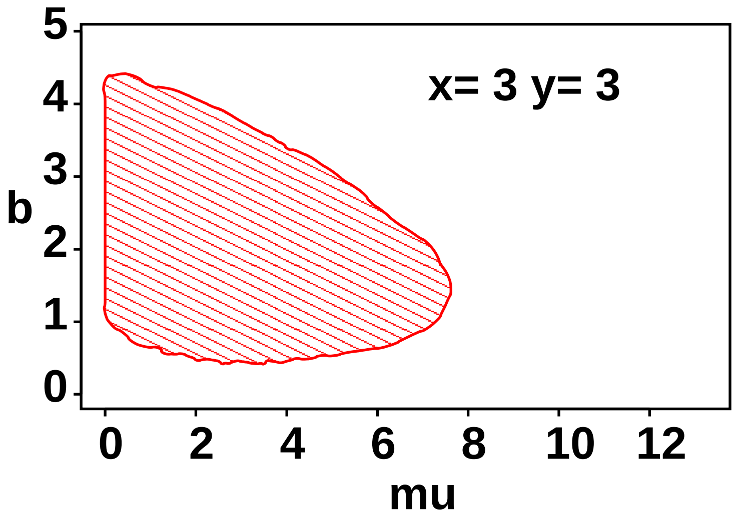
## A Subtlety, Our Tentative Solution

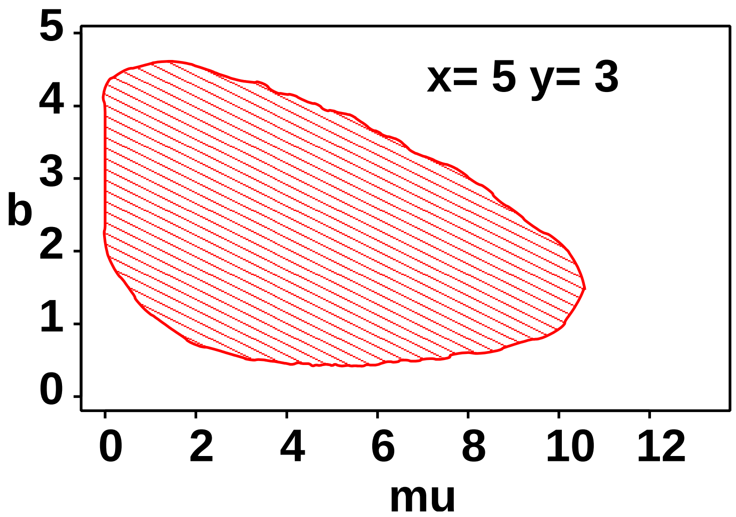



Gary Feldman 13 Fermilab Workshop

For $5\sigma$ discovery and $n \sim 100$ we don't care about $\pm 1$ over-coverage

## A Source of Confusion




**Rolke & Lopez** in their profile likelihood paper (hep-ph/0005187) first **start with a 2-dim Neyman construction using the likelihood as the ordering rule** (instead of likelihood ratio).

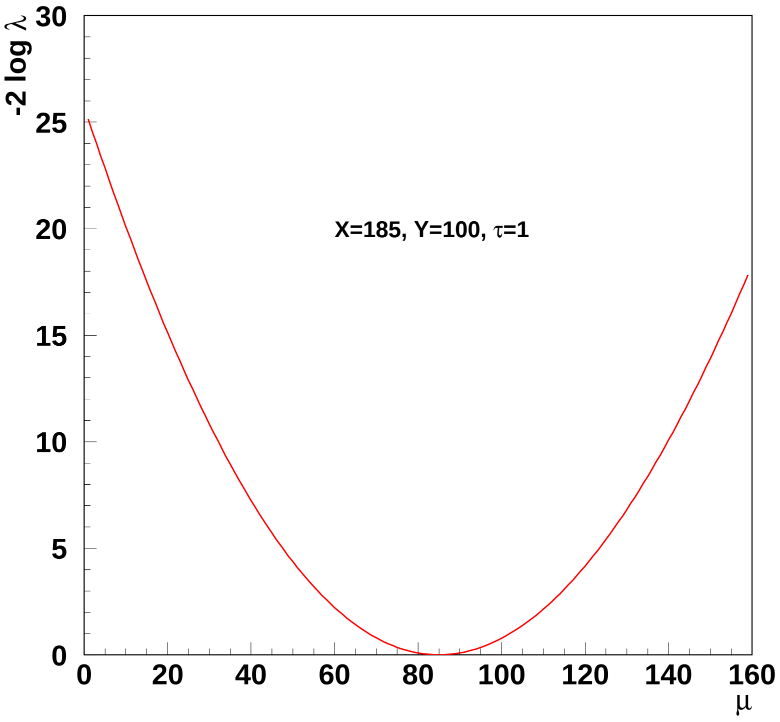
Then they go on to use the profile likelihood directly. **They don't actually use this construction**, except to fix pathological cases in the profile likelihood intervals.

**This confused Feldman & Cousins** at Fermilab workshop because they saw both a construction and the $\hat{b}$ equations and thought they were doing the same thing.

I asked `TRolke` to provide $5\sigma$ interval for $x = 165, y = 100, \tau = 1$, but I gave up after 5 hours

For the case: $L(X,Y|\mu,b) = Pois(X|\mu+b)\cdot Pois(Y|\tau b)$ the profile likelihood is for $\mu_0$ is given by $L(X,Y|\mu_0,\hat{\hat{b}}(\mu_0))$

$$\hat{\hat{b}}(\mu_0) = \frac{x+y-(1+\tau)\mu_0+\sqrt{(x+y-(1+\tau)\mu_0)^2+4(1+\tau)y\mu_0}}{2(1+\tau)}$$




The relevant likelihood ratio is then:

$$\lambda(\mu_0|x,y) = \frac{L(x,y|\mu_0,\hat{\hat{b}}(\mu)_0)}{L(x,y|\hat{\mu},\hat{b})}$$

The convergence of $-2\ln\lambda \approx \chi^2$ allows for approximate confidence intervals

In my tests with $y=100$ & $\tau=1$, one needs $x=185$ before $\mu=0$ is excluded at $5\sigma$. Minor under-coverage (4.8$\sigma$).

## Motivated by Louis, I took on the "double hats" for hypothesis testing

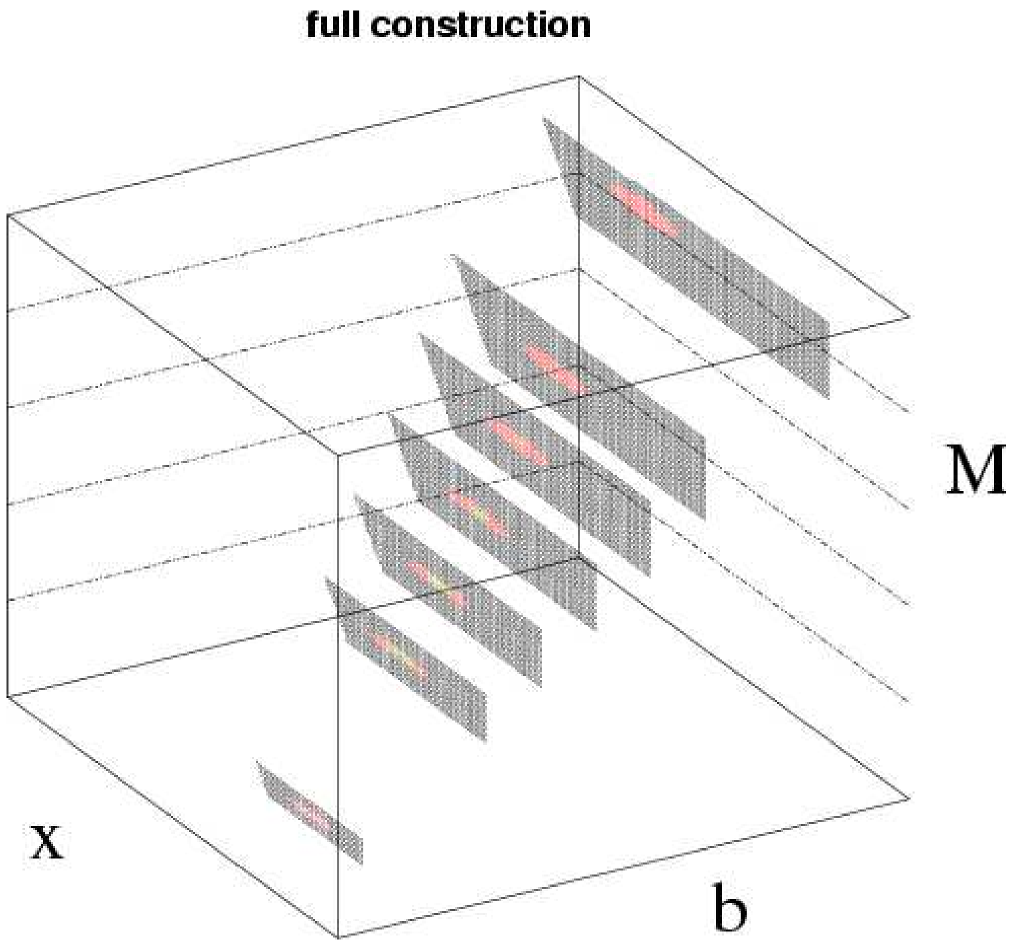
Now the parameter of interest is discrete, but I did the construction for all values of the nuisance parameter.

I used profile likelihood ratio as the ordering rule (which is independent of $b$)

$$l = \frac{L(x, y|H_0, \hat{\hat{b}})}{L(x, y|H_1, \hat{b})}$$

Ran into problem of similar tests. The tests are very similar over most of the range, but some irrelevant values of the nuisance parameters can cause problems.

Just put a window in $y$ around $L(x, y|H_0, b)$



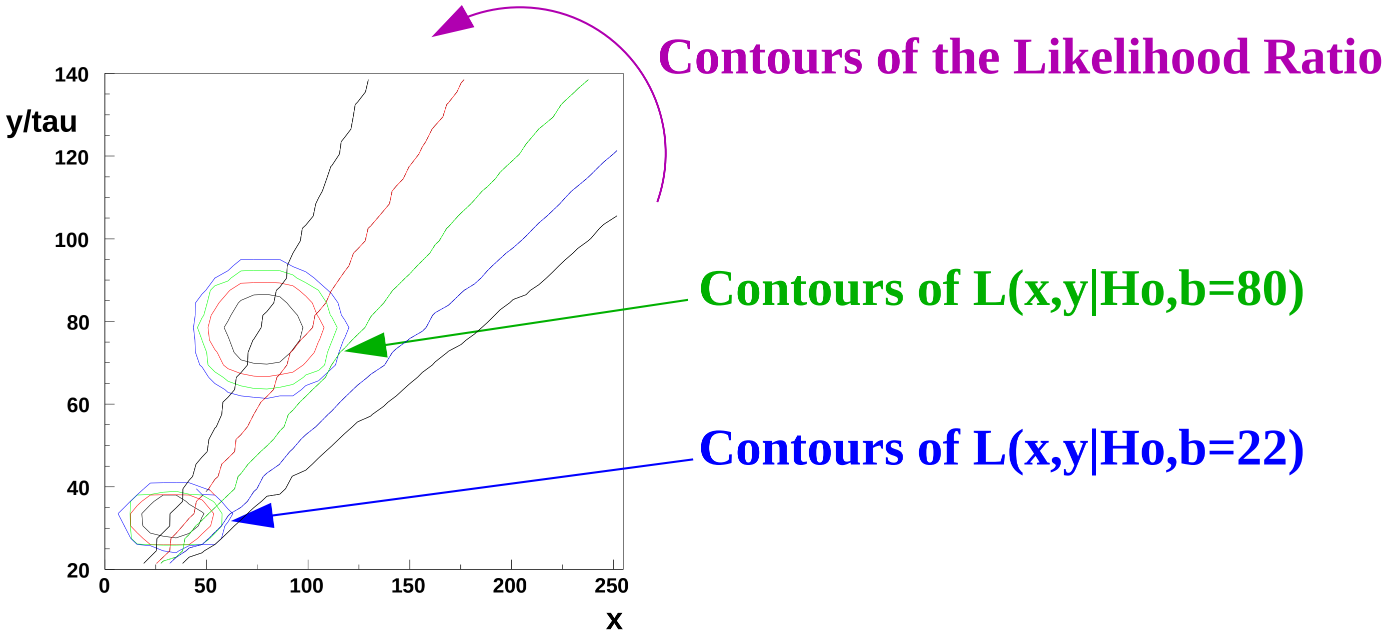

from PhyStat2003

Note Likelihood Ratio is independent of $b$

$$l = \frac{L(x,y|H_0,\hat{\hat{b}})}{L(x,y|H_1,\hat{b})}$$

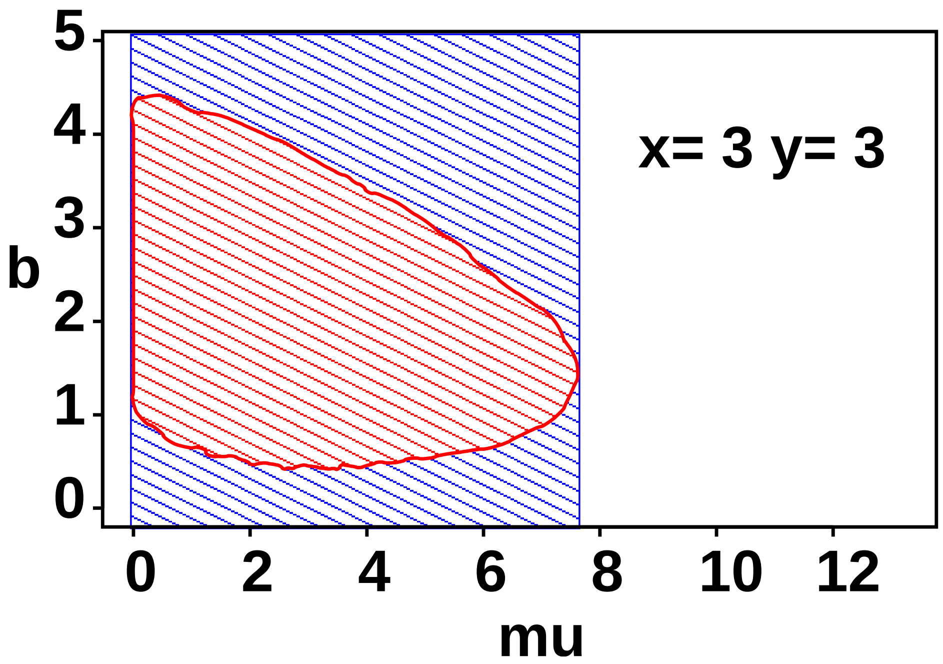
Similar Tests $\Rightarrow l_\alpha$ same for all $b$

Problem arises for $b << y/\tau$ where $l_\alpha$ needs to be smaller than for $b \approx y/\tau$




(See G. Punzi's talk later this week)

## Similar Tests & Projecting Out Nuisance Parameters




Rolke & Lopez made the entire construction including the nuisance parameter $b$

Could just project out $b$.

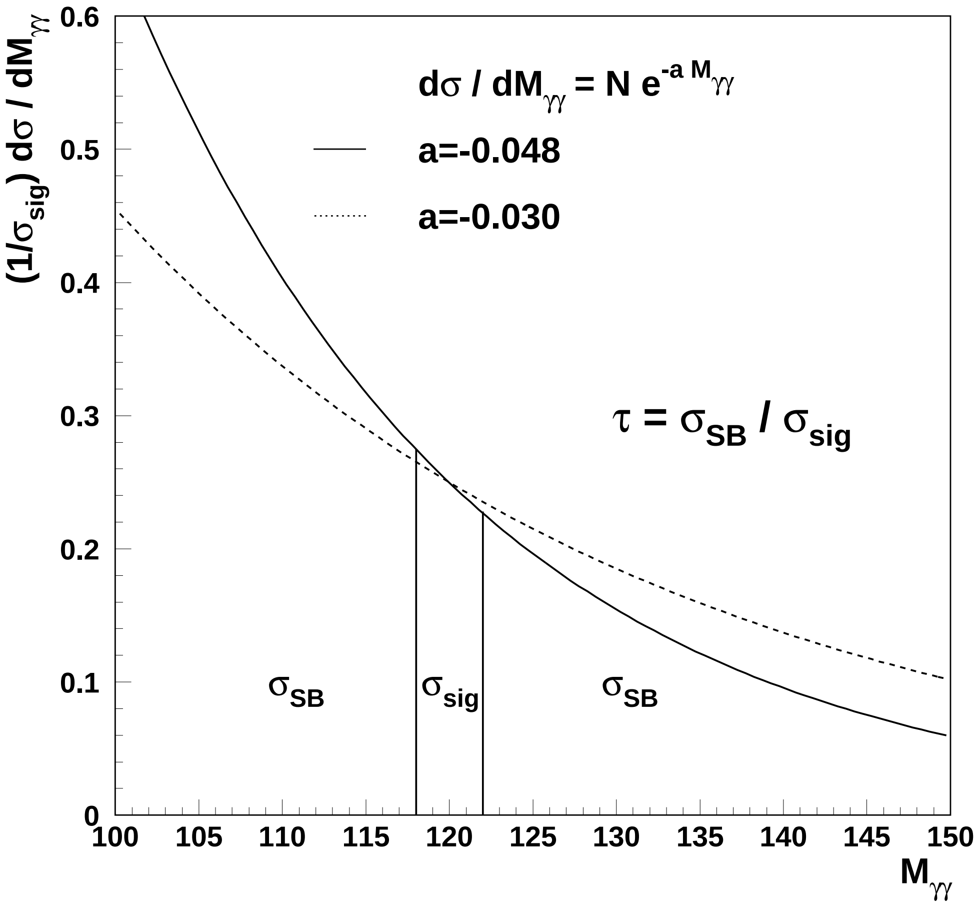
Common complaint is that this causes over-coverage (for some values of $b$)

In the case $\mu = 0$ is particularly important ($H_0$), we don't want to reject it if **any** value of $b$ can explain the data. (true for other nuisance parameters related to the null)

So the real challenge for hypothesis testing is providing approximately similar tests

*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 it always permits this elimination.* – A.W.F. Edwards

In this example, the ancillary measurement $Y$ is not a Poisson process, but the result of a fit to the sideband with some shape.




We used $d\sigma/dM_{\gamma\gamma} = Ne^{-aM_{\gamma\gamma}}$

We used a Toy MC to generate experiments with $\overline{N_{sig}} = 16000$ ($\approx 30\text{fb}^{-1}$).

We sampled the range $100 < M_{\gamma\gamma} < 150$ (about 200K events/exper).

We varied the exponent $a$ in the range $[-.048, -.030]$, generating nearly 1M experiments per exponent tested.

# *The $H \to \gamma\gamma$ Example (cont'd)*

The distribution of the expected background is very Gaussian.

A good guess for error is $\delta b \approx \sqrt{N_{SB}}/\tau$

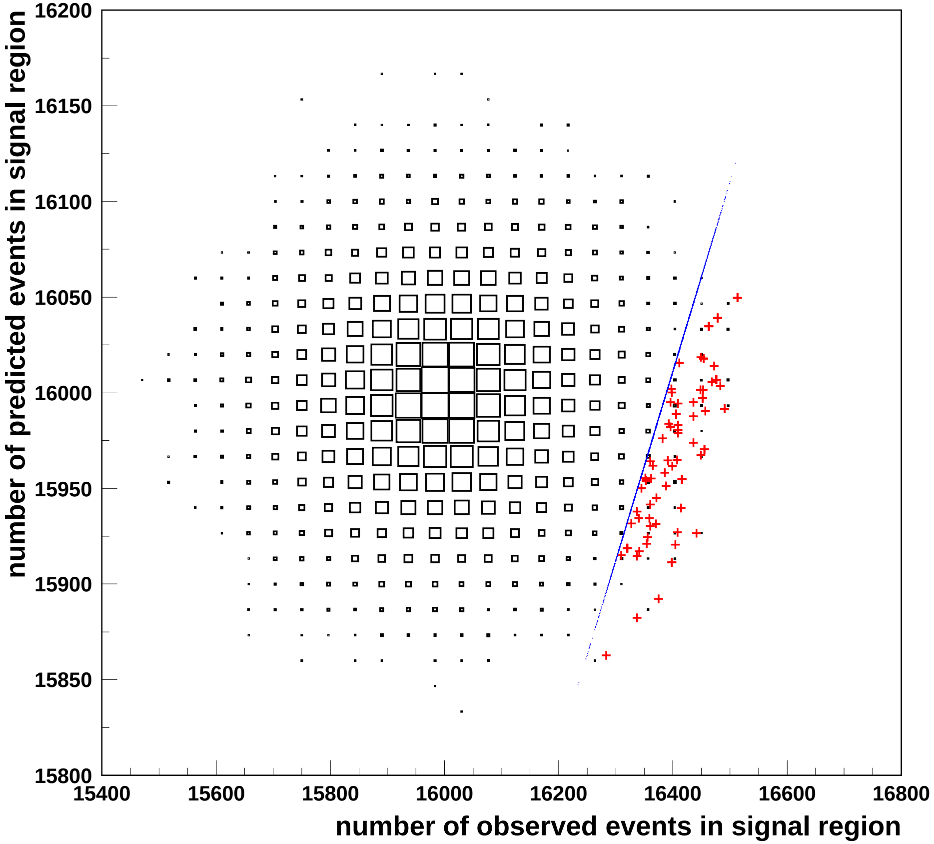
ATLAS often used what Linnemann called $Z_{5'}$ for expected significance

$$N\sigma = Z_{5'} = \frac{s}{\sqrt{b + (\delta b)^2}} = \underbrace{\frac{s}{\sqrt{b}\sqrt{1 + 1/\tau}}}$$

Equivalently, can define

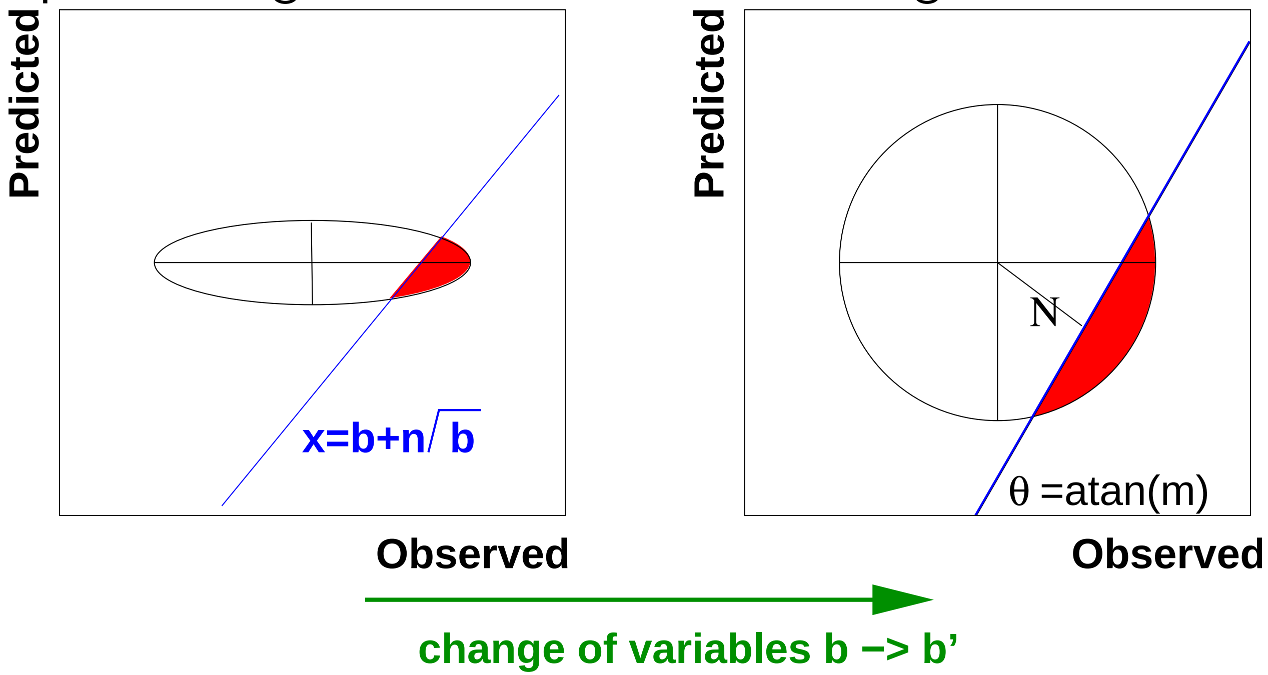
$$x_{crit}(b) = b + 5\sqrt{1 + 1/\tau}\sqrt{b}$$

Stathes Paganis and I showed the coverage of this method is fine




Note, $x_{crit}(b)$ is very linear for large $b$

Realizing that the form of the critical region is $x_{crit}(b) = b + n\sqrt{b}$ and that the boundary is very nearly linear around $b_0$, one can avoid using any ordering rule and just find the value of $n$ that gives the proper coverage and similar tests in the region of the measurement.

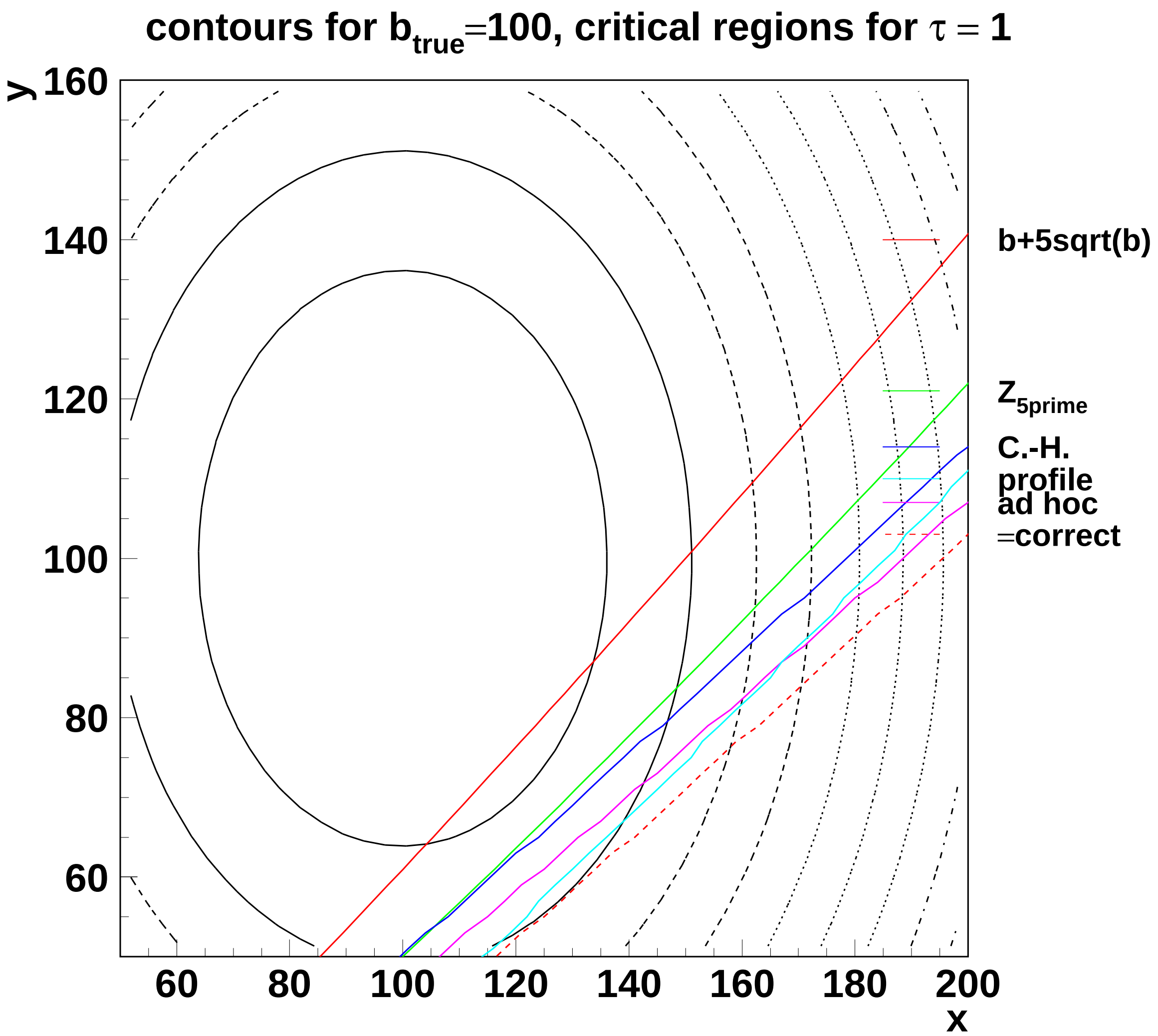



With a little geometry, we can find the number $n$ needed to get an $N\sigma$ test.

$$x_{crit}(b) = b + N\sqrt{1 + 1/\tau m^2}\sqrt{b} \quad \text{where} \quad m = \left(1 + \frac{N}{2\sqrt{b_0}}\right)^{-1}$$

The $m^2$ factor can be seen as a correction to the $Z_{5'}$ and $Z_N$ (Cousins-Highland) results

# *Comparison of Critical Regions for $5\sigma$*




(ovals indicate contours of true pdf)

Clearly the background uncertainty needs to be incorporated

The Cousins-Highland, it's approximation $Z_{5'}$, the profile likelihood (minuit), and the *ad hoc* critical region have moderate to significant under-coverage

(During original presentation, I had a mistake for the profile likelihood interval... it seemed to drastically over-cover. Corrected in this version)

For $y = 100$, the "reasonable methods" give a range of $x_{crit} \in [170, 188]$ and all undercover

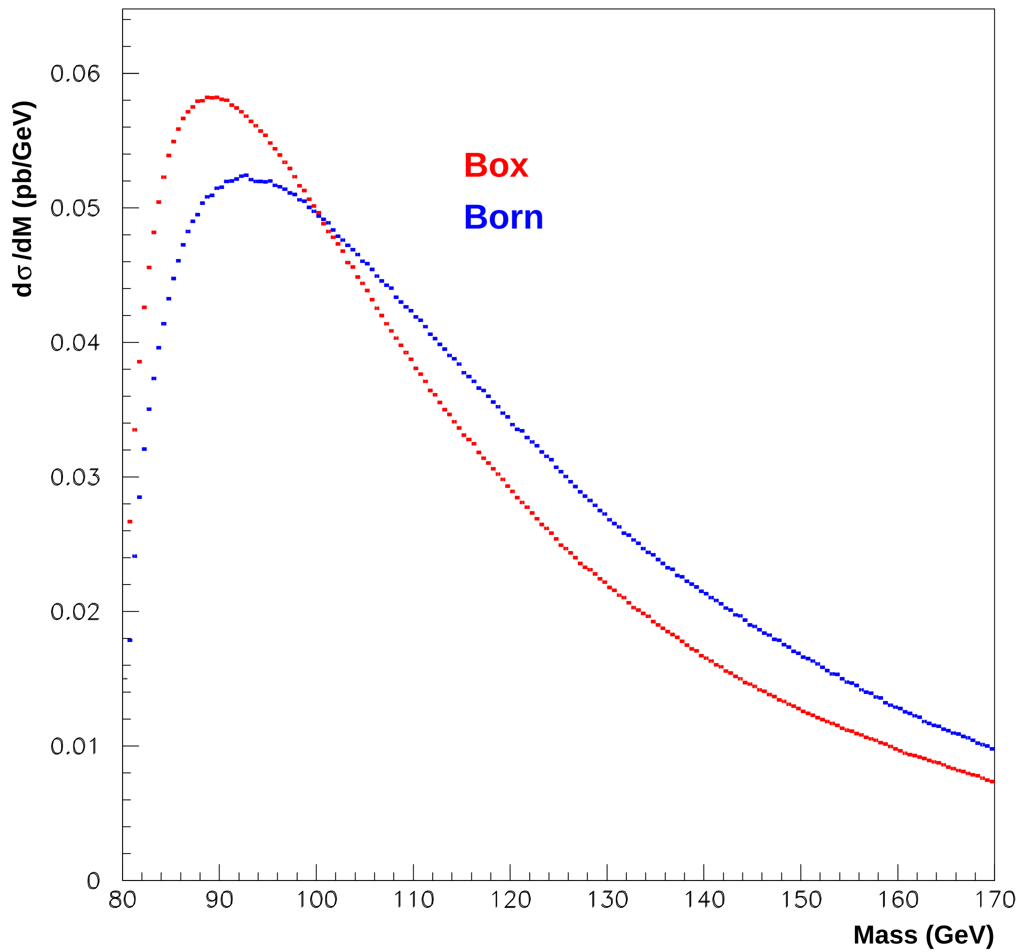
At PhyStat2003, Sinervo provided a classification of systematic errors.

The background uncertainty discussed so far is another statistical error (Class I)

In the $H \to \gamma\gamma$ example, there is also uncertainty on the shape of the continuum $M_{\gamma\gamma}$ spectrum.

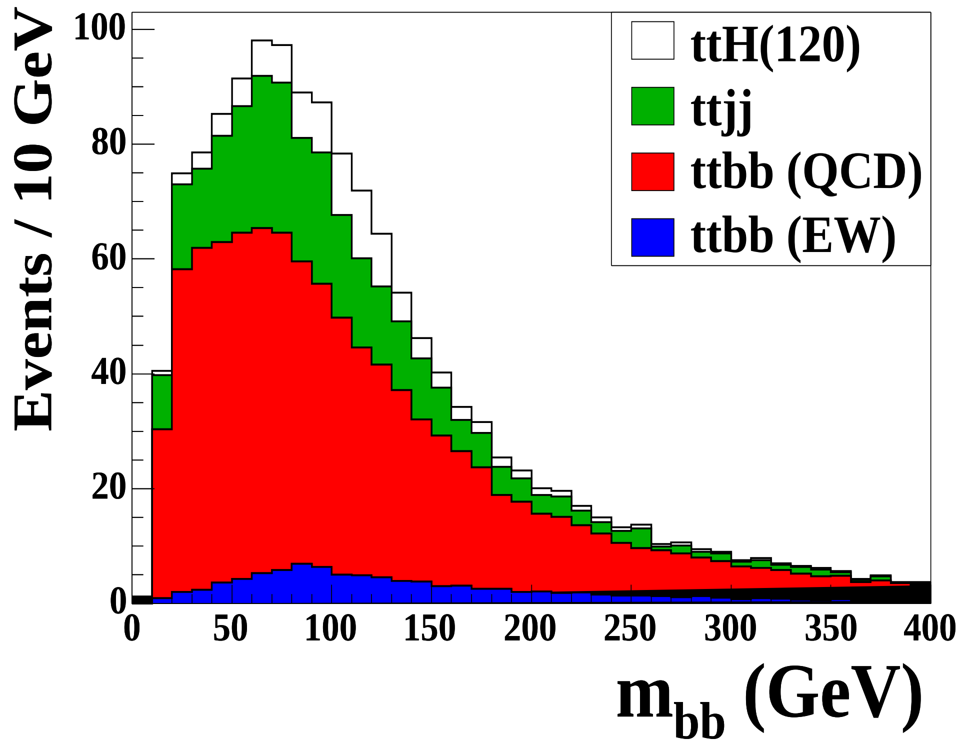
These shape uncertainties impact the background prediction from the sideband, and do not scale like statistical errors (Class II)

It seems plausible to extend the nuisance parameters to include the shape of the background. Similar to uncertainty on $\tau$.




Two plausible shapes for the continuum $M_{\gamma\gamma}$ spectrum.

J. Cammin's thesis on $ttH(H \to bb)$ in ATLAS




This is (was) one of the few powerful channels near the LEP limit

Combinatorial background is challenging with $4b$-jets and $\geq$ 6 jets total

Signal efficiency goes like $\epsilon_b^4$

Signal & bkgnd. have similar shape

Estimating $ttjj$ and $ttbb$ background from data difficult, large class II systematics

Neglecting systematics, expected significance $\approx 5\sigma$

Including 10% class II sytematic, it appears this channel will never reach $5\sigma$

Rolke & Lopez at PhyStat2003 considered three experimental procedures

- fixed mass, fixed width
- variable mass, fixed width
- variable mass, variable width

They illustrate the well-known problem that a sliding mass window will increase the rate of Type I error.

Yong-Sheng Gao at Southern Methodist University in Dallas has put out a note (hep-ex/0310011).

He is trying to address the dilution of significance in a sliding mass window approach

His suggestion is to do a maximum likelihood fit on the masses, cross-sections, and other parameters of the model, and then calculate the significance only at that point.

Somehow similar to the profile likelihood construction

It is of some comfort that more physicists seem to agree with the statement:

"Use whatever method you like, but check its coverage properties"

In common cases the coverage study can be used as justification for the use of an approximate method. (eg. Rolke, Lopez, Conrad, & Tegenfeldt)

However, in less common cases (*i.e.* the inevitable combination of various Higgs channels with systematic errors and discriminating variables in a few years time) this mantra is basically irrelevant.

In fact, the ability to do the coverage study implies that one can do the full construction, so why not just do it?

## *What looks promising?*

Unless we find a fairly general purpose asymptotic method that provides reasonable coverage for $5\sigma$ in the presence of large systematic errors, we should probably employ numerical methods roughly related to the Neyman construction

The profile construction looks promising: I don't know of any disastrous cases

For hypothesis testing, it seems that the proper thing to do is project-out the nuisance parameters associated with null hypothesis

If we move to the full construction (or in that direction), then the focus should be on roughly similar tests even if they are not "the most powerful" in some sense.

# Beyond the Standard Model

(Desert)

We are interested in testing for evidence of many diverse models with many model parameters and diverse phenomenology

- unconstrained MSSM with $\sim 105$ parameters
- mSUGRA has 4parameters +1sign
- little Higgs with masses, couplings, mixing angles
- black holes, gravitons, excited leptons, etc.

When the phenomenology is diverse

- no single observable will be sensitive to all model points.
- thus, we have the tradeoff: generality vs. power

One approach is to use benchmark points

- good for assessing sensitivity
- good for setting conservative limits
- irrelevant for determining model parameters from data

## Inclusive or "Quasi-Model Independent" Searches:

- look at distribution of some observable like $M_{eff}$
- used in ATLAS TDR, at H1 (hep-ex/0408044), at DØ (hep-ex/0006011)
- con: may not be sensitive, neglects a lot of discriminating power

## Dedicated Searches:

- look for evidence of some exclusive signature
- some analyses are very detailed (e.g. Higgs)
- con: time consuming, specific to a particular model

## "Recycled" and Composite Searches:

- many SUSY Higgs searches just use SM cuts and scale $\sigma BR$ (*i.e.* cuts are not optimal)
- combination between channels is powerful if done properly (*e.g.* consistent assumptions, correlations, etc.)

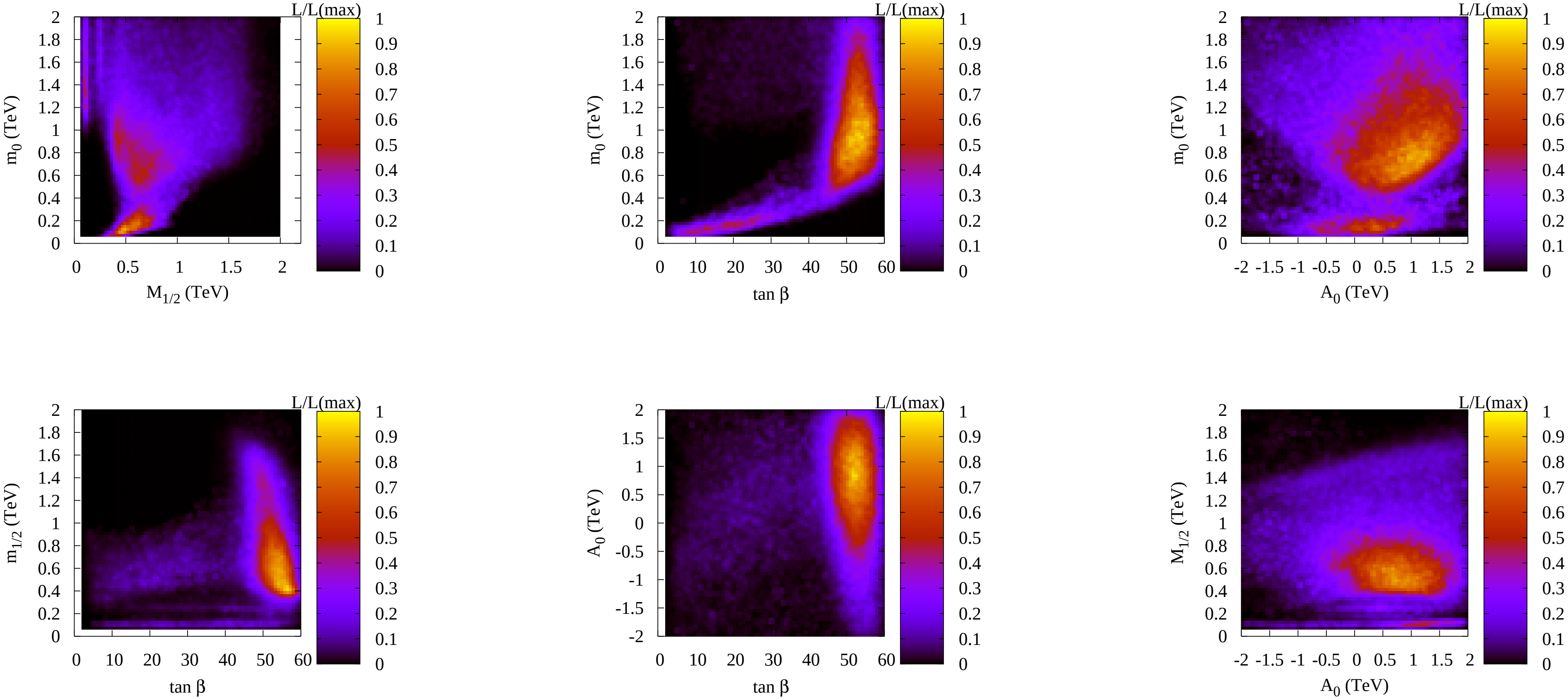
For much of Beyond Standard Model physics (*e.g.* SUSY) discovery is not the hard (interesting) part.

What we are more interested in is

- testing that a model gives an accurate description of the data
- sorting through models with fundamentally different theoretical motivation
- determining a model's parameters

Parameter Determination

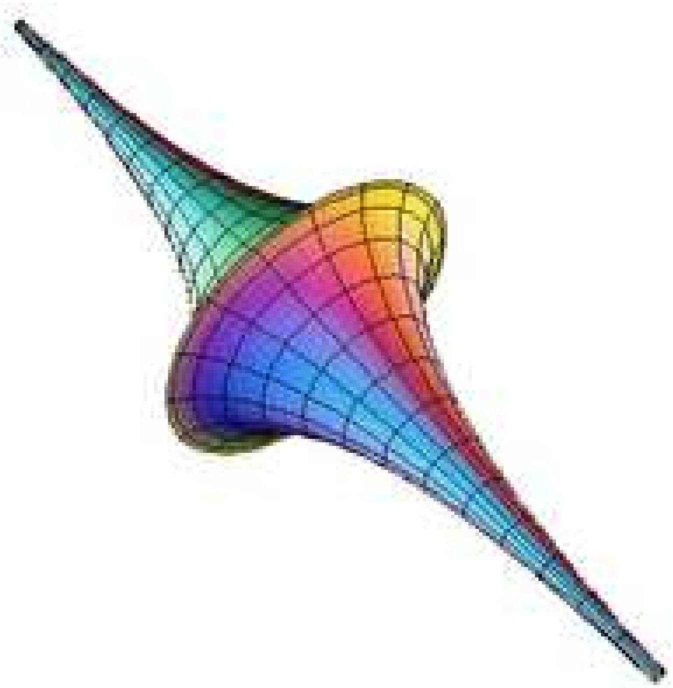
- SFITTER (Lafaye, Plehn, Zerwas hep-ph/0404282):
  - observables: mass spectrum,branching ratios, cross sections
  - fast grid search and/or detailed MINUIT fit
- Allanach (Cambridge) has employed a Markov Chain Monte Carlo
  - See hep-ph/0507283
  - sophisticated data analysis like this usually done by experimentalists
  - and most importantly...beautiful plots




Likelihood maps of mSUGRA parameter space. The graphs show the likelihood distributions sampled from 7d parameter space and marginalized down to two. The likelihood (relative to the likelihood in the highest bin) is displayed by reference to the bar on the right hand side of each plot. Projection Pursuit and related ideas could help here.

Amari considered the *Fisher Information Matrix* $g_{ij}$ *as a metric on a Manifold* $M$ *parametrized by* $\alpha$*:*

$$g_{ij}(\alpha) = \int dx f_\alpha(x) \left[\frac{\partial \log f_\alpha(x)}{\partial \alpha_i}\right] \left[\frac{\partial \log f_\alpha(x)}{\partial \alpha_j}\right]$$



Example:

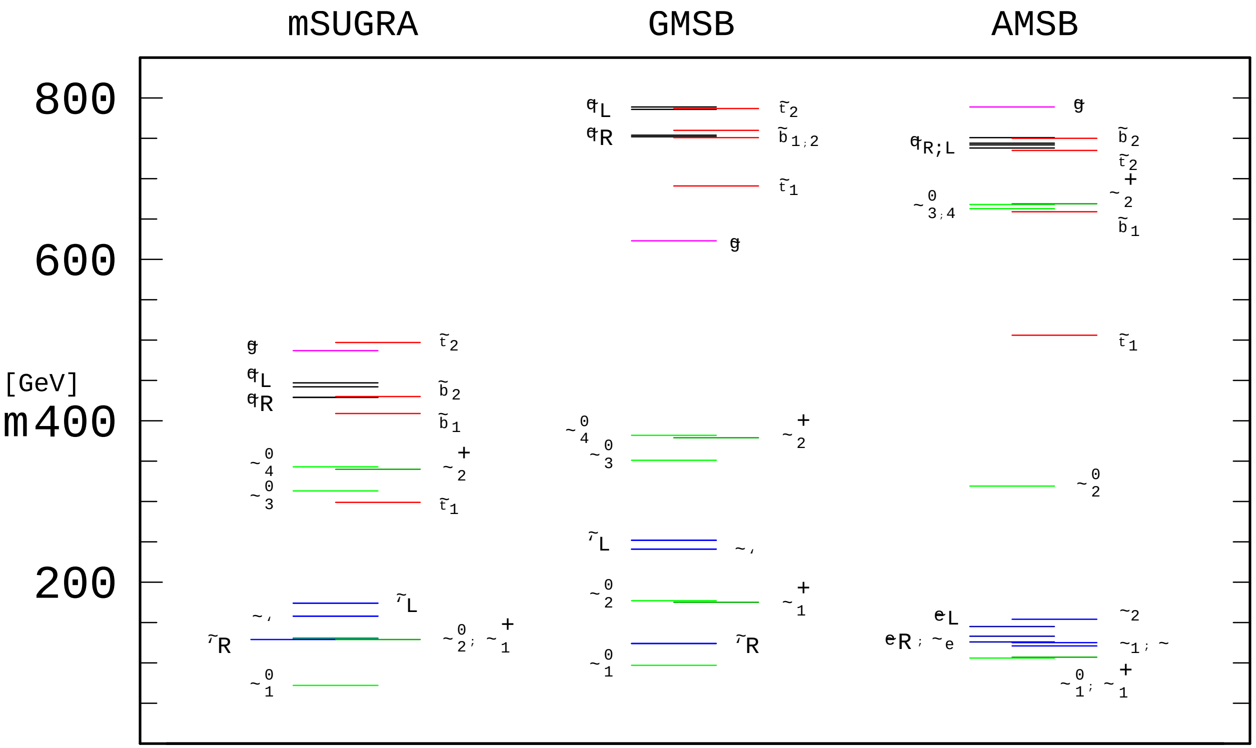
Consider Gaussians $G(x;\mu,\sigma)$ as a 2-d Manifold parametrized by $\alpha = (\mu,\sigma)$

the geometry is isotropic and negatively curved

*Natural Learning Rules* correspond to *geodesics* on the Manifold $M$.

Can lead to exponentially faster rates of convergence!

# *Information Geometry of MSSM*




We could try to use Information Geometry to improve how we sample the model space

An example use of Information Geometry for the MSSM:

- $\alpha =$ 105 model parameters
- $x =$ measured mass spectrum
- $f_\alpha(x) =$ probability to measure that spectrum given model parameters

At PhyStat2003, Brad Efron introduced the curious James-Stein estimator

$$\hat{\mu}_i = \bar{\mu}_i \left(1 - \frac{p-2}{||\bar{\mu}||}\right)$$

It can be shown that the mean squared error of $\hat{\mu}$ is always smaller than the maximum likelihood estimator $\bar{\mu}$!

The improvement in mean squared error depends on the "prior prior" of the $\mu_i$. James-Stein is best for Gaussian "prior prior", though there are generalizations.

For techniques like SFitter, the input to the parameter determination is $\bar{\mu}$ and $p \approx 15$.

...but $\hat{\mu}_i$ is biased for each $i$

Can we make use of this in physics? Can we understand it? For the "prior prior" of SUSY masses, what is the equivalent estimator?

# Conclusions

# &

# Final Remarks

## *Where Statisticians can help*

Can there be progress to improve asymptotic techniques with the focus on the description of tails?

At what value of $N$ do the improved profile likelihood methods help? Are they appropriate for our applications?

What is the generalization of the James-Stein estimator for the expected SUSY mass values? Should it be used?

## *Conclusions*

Proper treatment of systematics is vital for new particle searches at the LHC

We basically understand the relevant issues and the formalism, but our existing tools are not sufficient

We need a very general purpose method that can reliably assess $5\sigma$ tails

- Probably need numerical methods guided by asymptotic results

The final significance calculation will be so complicated, coverage studies for approximate methods may be irrelevant

# Backup Slides


Archbishop of Canterbury Thomas Cranmer
born: 1489 executed: 21 March 1556
author of the "Book of Common Prayer"

Two centuries later, (when this Book had become an official prayer book of the Church of England) Thomas Bayes was a Non-conformist minister (Presbyterian) who refused to use Cranmer's book.

Therefore, I will be talking about Frequentist Methods.

## *References*

Gary Feldman. "Multiple Measurements and Parameters in the Unified Approach" Workshop on Confidence Limits, Fermilab, March 28, 2000.

Stuart, Ord, Arnold. "Kendall's Advanced Theory of Statistics" Vol. 2A *Classical Inference & the Linear Model.*

(note: the chapters move around with different volumes.
Chapter 24 of vol 2 = Chapter 22 of vol 2A.)

(also see Section 21.20 "The choice of most powerful similar regions")

(also see Section 20.1 - 20.11 For Neyman Pearson Theory)

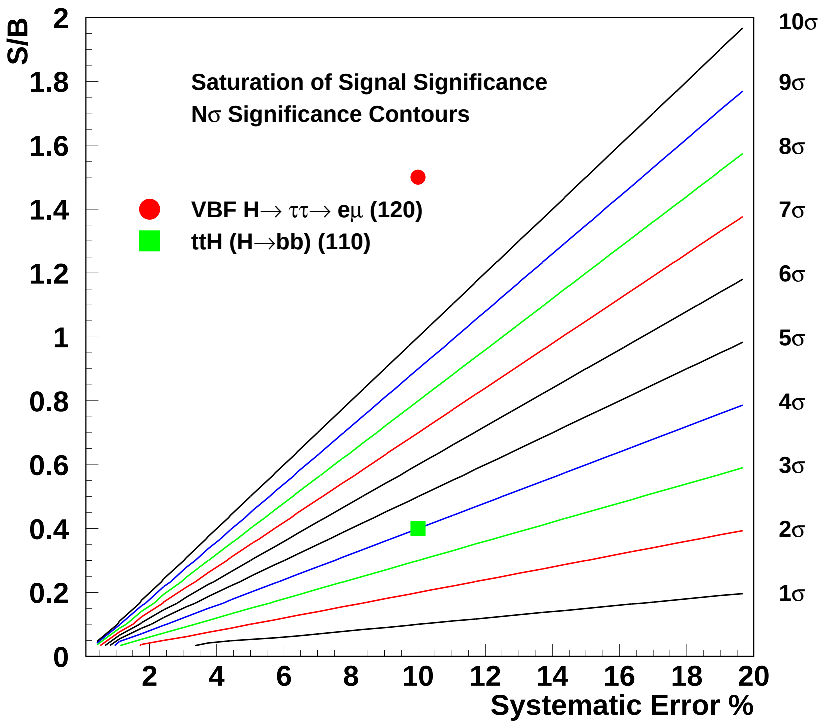
(also see Table 20.1 For "Relationships between hypothesis testing and interval estimation")

Background determination from sidebands carries two sources of error:

- statistical error from sideband measurement

- systematic on extrapolation from sideband to signal-like area (shape systematic)

The shape systematic does not (necessarily) reduce with increased luminosity



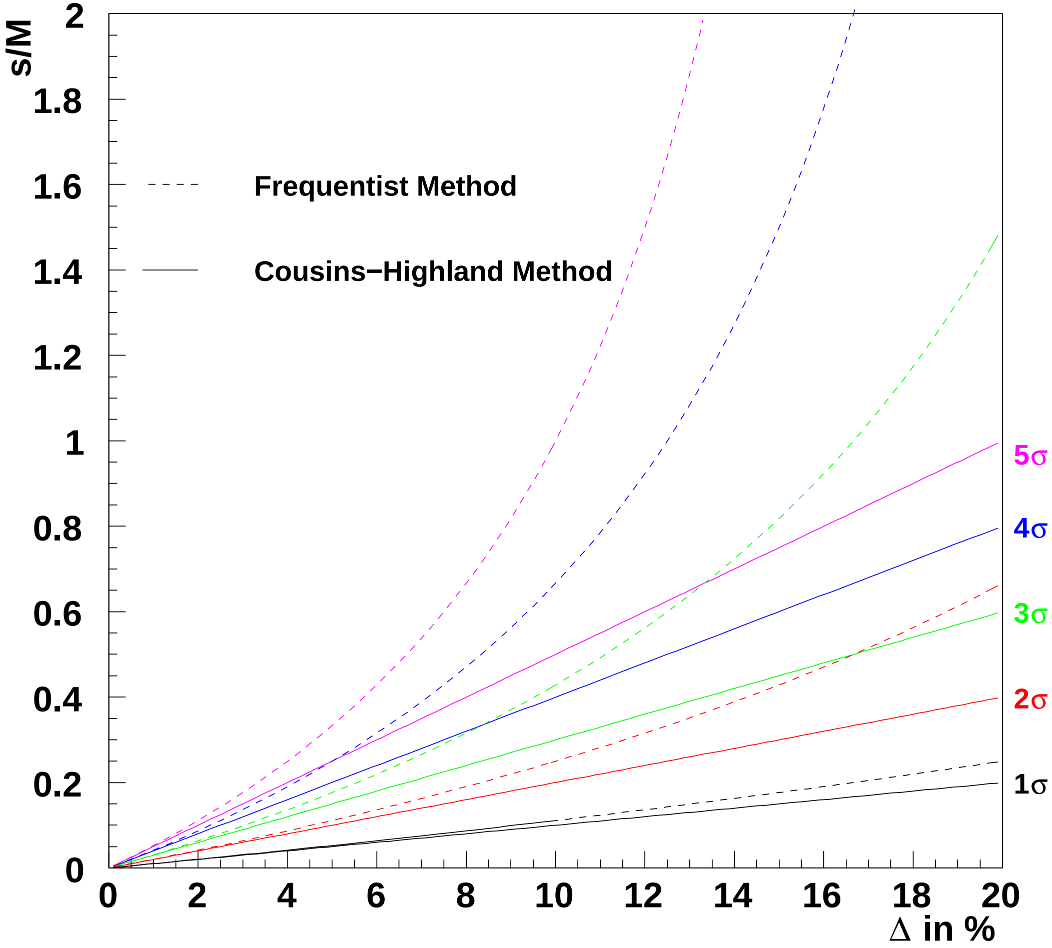

Normal significance measure $s/\sqrt{b}$ is replaced by $s/\sqrt{b(1+b\Delta^2)}$

If $s/b$ is fixed as we increase luminosity, the expected significance saturates:

$$\sigma_\infty = \frac{s/b}{\Delta_{shape}}$$

With its low $S/B$ and 10% shape ystematic, $ttH(\to bb)$ can't get to $5\sigma$ even with $L \to \infty$

# *Backup: Limiting Behavior of Proposed Method*




PRELIMINARY

With Frequentist Technique approximate limiting behavior given by:

If $\Delta = \delta M/M$ and $s/M$ fixed as we increase luminosity, the significance saturates:

$$\sigma_\infty = \frac{s}{\Delta(s+M)} \tag{1}$$

Note for Gaussian $L(M|b)$: limited by $\sigma_\infty < 1/\Delta$

In 1928-1938 Neyman & Pearson developed a theory in which one must consider competing Hypotheses:

- the Null Hypothesis $H_0$ (background only)
- the Alternate Hypothesis $H_1$ (signal-plus-background)

Given some probability that we wrongly reject the Null Hypothesis

$$\alpha = P(x \notin W | H_0)$$

Find the region $W$ such that we minimize the probability of wrongly accepting $H_0$ (when $H_1$ is true)

$$\beta = P(x \in W | H_1)$$

The region $W$ that minimizes the probability of wrongly accepting the $H_0$ is just a contour of the Likelihood Ratio:

$$\frac{L(x|H_0)}{L(x|H_1)} > k_\alpha$$

Power $=$ probability of rejecting false $H_0$

This is the basis for the Unified Method's Ordering Rule

Have "Compound Hypothesis" $L(x|\mu)$ where $\mu$ is an unknown physics parameter and $\mu_t$ is the "true value"

No notion of $L(\mu)$ – that's Bayesian!

Frequentist ask for a "Confidence Interval":
an interval of $\mu$ which will contain $\mu_t$ with probability $1 - \alpha$.

The Interval is not unique!
Power=1-probability to accept false $\mu$

The *ordering rule* defines the interval

## Feldman & Cousins "Unified Approach" looks like this:

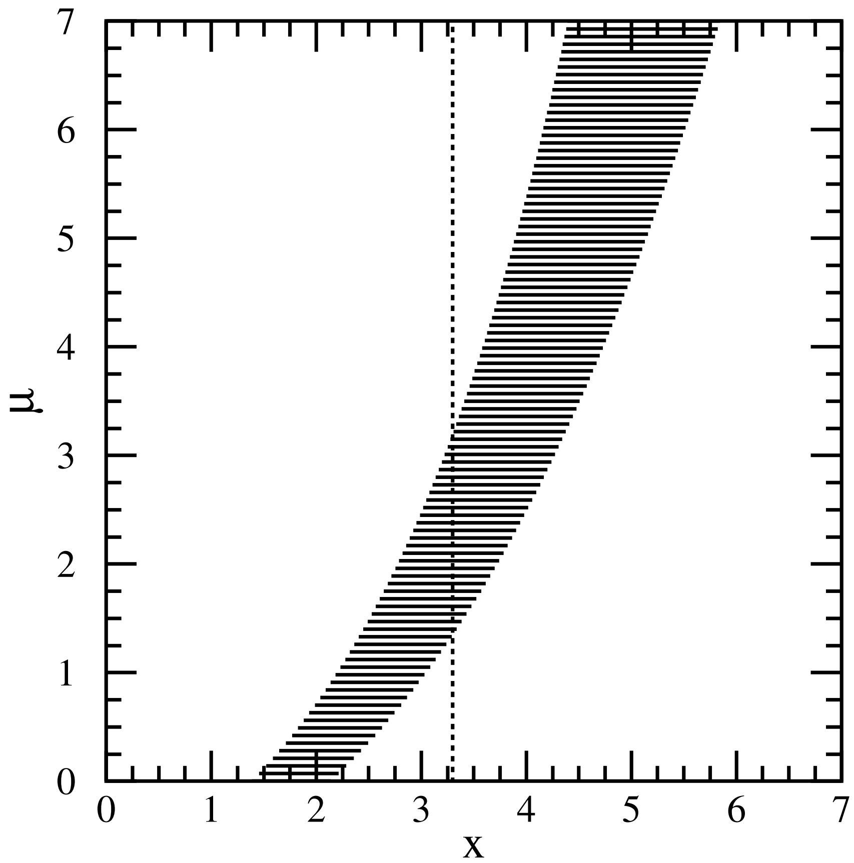
Neyman Construction

- For each $\mu$: find region $R_\mu$ with probability $1-\alpha$
- Confidence Interval includes all $\mu$ consistent with observation at $x_0$

Ordering Rule specifies what region

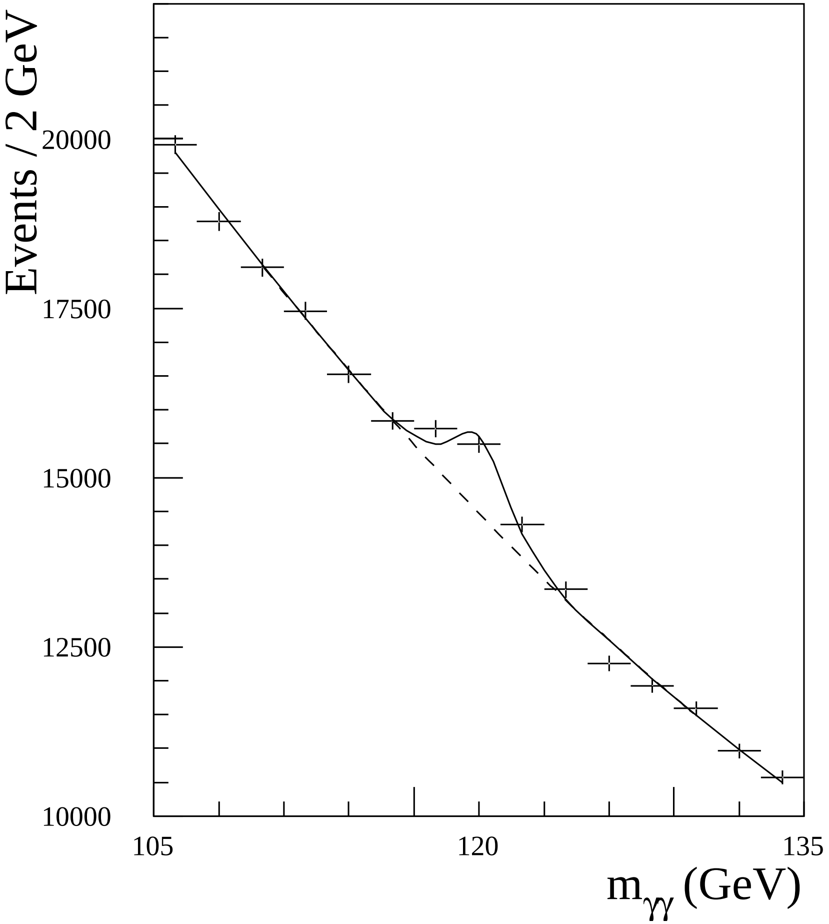
F-C ordering rule is the Likelihood Ratio

$$R_\mu = \left\{ x \;\middle|\; \frac{L(x|\mu)}{L(x|\mu_{\rm best})} > k_\alpha \right\}$$




**The F-C ordering rule follows naturally from Neyman-Pearson Lemma**

In our Scientific Note for $30\text{fb}^{-1}$, we expect $s = 385$, $b = 14190$




Without systematic uncertainty:
$N\sigma = s/\sqrt{b} = 3.2\sigma$

Using the binning in our studies, $\tau = 12.5$:
$N\sigma = \frac{s}{\sqrt{b}\sqrt{1+1/\tau}} = 3.1\sigma$

Using the binning in the TDR plot, $\tau = 4$:
$N\sigma = \frac{s}{\sqrt{b}\sqrt{1+1/\tau}} = 2.9\sigma$

Same result with Frequentist approach.