Is there evidence for a peak in this data?
Is there evidence for a peak in this data?

“Observation of an Exotic $S=+1$ Baryon in Exclusive Photoproduction from the Deuteron”

“The statistical significance of the peak is $5.2 \pm 0.6 \sigma$”
Is there evidence for a peak in this data?

“Observation of an Exotic S=+1 Baryon in Exclusive Photoproduction from the Deuteron”
“The statistical significance of the peak is 5.2 ± 0.6 σ”

“A Bayesian analysis of pentaquark signals from CLAS data”
“The In(RE) value for g2a (-0.408) indicates weak evidence in favour of the data model without a peak in the spectrum.”

Statistical Issues in Searches for New Physics

Louis Lyons
Imperial College, London
and
Oxford
Theme: Using data to make judgements about H1 (New Physics) versus H0 (S.M. with nothing new)

Why?
Experiments are expensive and time-consuming
so
Worth investing effort in statistical analysis
→ better information from data

Topics:
  Blind Analysis
  LEE = Look Elsewhere Effect
  Why 5σ for discovery?
  Significance
  P(A|B) ≠ P(B|A)
  Meaning of p-values
  Wilks’ Theorem
  Background Systematics
  Coverage
  Example of misleading inference
  p₀ v p₁ plots
  Higgs search: Discovery and spin
  (N.B. Several of these topics have no unique solutions from Statisticians)

Conclusions

Extended version of talk at LHCP2014 in New York, and CERN Seminar 2015
Choosing between 2 hypotheses

Hypothesis testing: New particle or statistical fluctuation?

\[ H_0 = b \quad H_1 = b + s \]
Choosing between 2 hypotheses

Possible methods:

$\Delta \chi^2$

$p$-value of statistic

$\ln L$–ratio

Bayesian:

Posterior odds

Bayes factor

Bayes information criterion (BIC)

Akaike ………

(AIC)

Minimise “cost”

See ‘Comparing two hypotheses’

With 2 hypotheses, each with own pdf, p-values are defined as tail areas, pointing in towards each other.
Procedure for choosing between 2 hypotheses

1) No sensitivity

2) Maybe

3) Easy separation

Procedure:

Obtain expected distributions for data statistic (e.g. $L$-ratio) for H0 and H1

Choose $\alpha$ (e.g. 95%, 3$\sigma$, 5$\sigma$) and CL for $p_1$ (e.g. 95%)

Given $b$, $\alpha$ determines $t_{\text{crit}}$

$b+s$ defines $\beta$. For $s > s_{\text{min}}$, separation of curves $\rightarrow$ discovery or excln

$1-\beta = \text{Power of test}$

Now data:

If $t_{\text{obs}} \geq t_{\text{crit}}$ (i.e. $p_0 \leq \alpha$), discovery at level $\alpha$

If $t_{\text{obs}} < t_{\text{crit}}$, no discovery. If $p_1 < 1-\text{CL}$, exclude H1
BLIND ANALYSES

Why blind analysis? Data statistic, selections, corrections, method

Methods of blinding
- Add random number to result *
- Study procedure with simulation only
- Look at only first fraction of data
- Keep the signal box closed
- Keep MC parameters hidden
- Keep unknown fraction visible for each bin

Disadvantages
- Takes longer time
- Usually not available for searches for unknown

After analysis is unblinded, don’t change anything unless ........

* Luis Alvarez suggestion re “discovery” of free quarks
Look Elsewhere Effect (LEE)

Prob of bgd fluctuation at that place = local p-value
Prob of bgd fluctuation ‘anywhere’ = global p-value

Global p > Local p

Where is ‘anywhere’?

a) Any location in this histogram in sensible range
b) Any location in this histogram
c) Also in histogram produced with different cuts, binning, etc.
d) Also in other plausible histograms for this analysis
e) Also in other searches in this PHYSICS group (e.g. SUSY at CMS)
f) In any search in this experiment (e.g. CMS)
g) In all CERN expts (e.g. LHC expts + NA62 + OPERA + ASACUSA + ....)
h) In all HEP expts
    etc.

d) relevant for graduate student doing analysis
f) relevant for experiment’s Spokesperson

INFORMAL CONSENSUS:
Quote local p, and global p according to a) above.
Explain which global p
Example of LEE: Stonehenge
12 is the number of constellations

6 is the number of ages (2160) we spend on each side of the galactic equator

18 number of breaths we take each minute or our life

Missing two large stones in top half. Should be 6 and 6

1 degree = 72 years
360 x 72 = 25,920

25,920 divided by 60 = 432
432 x 5 = 2,160

Should be 5 stones between each division on the Second ring.

If small stones = 432 years each then the half circle in the center would be 20 x 432 = 8640 years
8640 divided by 2160 = 4th time.

25,920 divided by 12 = 2160

25,920 divided by 6 = 4320

25,920 divided by 18 = 1440

Center Stone in Center Ring would be divided in half by sun rays when Earth in perfect balance. Nine on each side + 2 = 20.

30 Stones in Outer ring = 360 divided by 30 = 12

60 Stones in Second ring = 360 divided by 60 = 6

20 Stones in Center ring = 360 divided by 20 = 18

STONEHENGE

The Book of Truth
A New Perspective on the Hopi Creation Story
by Thomas O. Mills

STONEHENGE from a Hopi point of view.

Doesn’t make sense with today’s eastward direction.
Are alignments significant?

- Atkinson replied with his article "Moonshine on Stonehenge" in *Antiquity* in 1966, pointing out that some of the pits which ..... had used for his sight lines were more likely to have been natural depressions, and that he had allowed a margin of error of up to 2 degrees in his alignments. Atkinson found that the probability of so many alignments being visible from 165 points to be close to 0.5 rather than the "one in a million" possibility which ..... had claimed.

- ..... had been examining stone circles since the 1950s in search of astronomical alignments and the *megalithic yard*. It was not until 1973 that he turned his attention to Stonehenge. He chose to ignore alignments between features within the monument, considering them to be too close together to be reliable. He looked for landscape features that could have marked lunar and solar events. However, one of .....'s key sites, Peter's Mound, turned out to be a twentieth-century rubbish dump.
Why 5\(\sigma\) for Discovery?

Statisticians ridicule our belief in extreme tails (esp. for systematics)

Our reasons:

1) Past history (Many 3\(\sigma\) and 4\(\sigma\) effects have gone away)
2) LEE (see later)
3) Worries about underestimated systematics
4) Subconscious Bayes calculation

\[
\frac{p(H_1|x)}{p(H_0|x)} = \frac{p(x|H_1)}{p(x|H_0)} \times \frac{\pi(H_1)}{\pi(H_0)}
\]

Posterior     Likelihood   Priors
prob          ratio

“Extraordinary claims require extraordinary evidence”

N.B. Points 2), 3) and 4) are experiment-dependent

Alternative suggestion:

L.L. “Discovering the significance of 5\(\sigma\)” http://arxiv.org/abs/1310.1284
How many $\sigma$’s for discovery?

<table>
<thead>
<tr>
<th>SEARCH</th>
<th>SURPRISE</th>
<th>IMPACT</th>
<th>LEE</th>
<th>SYSTEMATICS</th>
<th>No. $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higgs search</td>
<td>Medium</td>
<td>Very high</td>
<td>M</td>
<td>Medium</td>
<td>5</td>
</tr>
<tr>
<td>Single top</td>
<td>No</td>
<td>Low</td>
<td>No</td>
<td>No</td>
<td>3</td>
</tr>
<tr>
<td>SUSY</td>
<td>Yes</td>
<td>Very high</td>
<td>Very large</td>
<td>Yes</td>
<td>7</td>
</tr>
<tr>
<td>$B_s$ oscillations</td>
<td>Medium/Low</td>
<td>Medium</td>
<td>$\Delta m$</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Neutrino osc</td>
<td>Medium</td>
<td>High</td>
<td>$\sin^2 2\theta, \Delta m^2$</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>$B_s \rightarrow \mu \mu$</td>
<td>No</td>
<td>Low/Medium</td>
<td>No</td>
<td>Medium</td>
<td>3</td>
</tr>
<tr>
<td>Pentaquark</td>
<td>Yes</td>
<td>High/V. high</td>
<td>M, decay mode</td>
<td>Medium</td>
<td>7</td>
</tr>
<tr>
<td>$(g-2)_\mu$ anom</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>$H$ spin $\neq 0$</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>Medium</td>
<td>5</td>
</tr>
<tr>
<td>4$^{th}$ gen q, l, $\nu$</td>
<td>Yes</td>
<td>High</td>
<td>M, mode</td>
<td>No</td>
<td>6</td>
</tr>
<tr>
<td>Dark energy</td>
<td>Yes</td>
<td>Very high</td>
<td>Strength</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>Grav Waves</td>
<td>No</td>
<td>High</td>
<td>Enormous</td>
<td>Yes</td>
<td>8</td>
</tr>
</tbody>
</table>

Suggestions to provoke discussion, rather than `delivered on Mt. Sinai’

Bob Cousins: “2 independent expts each with 3.5$\sigma$ better than one expt with 5$\sigma$”
Significance

Significance = \frac{S}{\sqrt{B}} \quad \text{or similar?}

Potential Problems:

• Uncertainty in B
• Non-Gaussian behaviour of Poisson, especially in tail
• Number of bins in histogram, no. of other histograms [LEE]
• Choice of cuts, bins (Blind analyses)

For future experiments:

• Optimising: Could give \( S = 0.1, B = 10^{-4}, \quad \frac{S}{\sqrt{B}} = 10 \)
$P(A | B) \neq P(B | A)$

Remind Lab or University media contact person that:

Prob[data, given H0] is very small
does not imply that
Prob[H0, given data] is also very small.

e.g. Prob{data | speed of $\nu \leq c$}= very small
does not imply
Prob{speed of $\nu \leq c$ | data} = very small
or
Prob{speed of $\nu > c$ | data} $\sim$ 1

Everyday situation, 2$^{nd}$ most convincing example:
Pack of playing cards

$p(\text{spade}|\text{king}) = 1/4$
p($\text{king}|\text{spade}$) = 1/13
\[ P(A | B) \neq P(B | A) \]

Remind Lab or University media contact person that:

Prob[data, given H0] is very small
does not imply that
Prob[H0, given data] is also very small.

e.g. Prob[data | speed of \( v \leq c \)] = very small
does not imply
Prob[speed of \( v \leq c \) | data] = very small
or
Prob[speed of \( v > c \) | data] \sim 1

Everyday example

\[ p(\text{pregnant}|\text{female}) \sim 3\% \]

\[ p(\text{female}|\text{pregnant}) >> 3\% \]
What p-values are (and are not)

Reject $H_0$ if $t > t_{\text{crit}}$ ($p < \alpha$)

$p$-value = prob that $t \geq t_{\text{obs}}$

Small $p \rightarrow$ data and theory have poor compatibility

Small $p$-value does NOT automatically imply that theory is unlikely

Bayes $\text{prob}(\text{Theory} | \text{data})$ related to $\text{prob}(\text{data} | \text{Theory}) = \text{Likelihood}$

by Bayes Th, including Bayesian prior

$p$-values are misunderstood. e.g. Anti-HEP jibe:

“Particle Physicists don’t know what they are doing, because half their $p < 0.05$ exclusions turn out to be wrong”

Demonstrates lack of understanding of $p$-values

[All results rejecting energy conservation with $p < \alpha = .05$ cut will turn out to be ‘wrong’]
Combining different p-values

Several results quote independent p-values for same effect:
\( p_1, p_2, p_3 \ldots \) e.g. 0.9, 0.001, 0.3 \ldots

What is combined significance? Not just \( p_1 \cdot p_2 \cdot p_3 \ldots \)

If 10 expts each have \( p \sim 0.5 \), product \( \sim 0.001 \) and is clearly **NOT** correct
combined \( p \)

\[
S = z \cdot \sum_{j=0}^{n-1} \frac{(-\ln z)^j}{j!}, \quad z = p_1 p_2 p_3 \ldots
\]

(e.g. For 2 measurements, \( S = z \cdot (1 - \ln z) \geq z \) )

Problems:

1) **Recipe is not unique** (Uniform dist in n-D hypercube \( \rightarrow \) uniform in 1-D)
2) **Formula is not associative**

Combining \( \{\{p_1 \text{ and } p_2\}\}, \text{ and then } p_3 \} \) gives different answer
from \( \{\{p_3 \text{ and } p_2\}, \text{ and then } p_1\} \), or all together
Due to different options for “more extreme than \( x_1, x_2, x_3 \)”.
3) Small \( p \)'s due to different discrepancies

******* Better to combine data ************
Wilks’ Theorem

Data = some distribution e.g. mass histogram
For H0 and H1, calculate best fit weighted sum of squares $S_0$ and $S_1$
Examples: 1) $H_0 =$ polynomial of degree 3
   $H_1 =$ polynomial of degree 5
2) $H_0 =$ background only
   $H_1 =$ bgd+peak with free $M_0$ and cross-section
3) $H_0 =$ normal neutrino hierarchy
   $H_1 =$ inverted hierarchy

If $H_0$ true, $S_0$ distributed as $\chi^2$ with ndf = $\nu_0$
If $H_1$ true, $S_1$ distributed as $\chi^2$ with ndf = $\nu_1$
If $H_0$ true, what is distribution of $\Delta S = S_0 - S_1$? Expect not large. Is it $\chi^2$?

Wilks’ Theorem: $\Delta S$ distributed as $\chi^2$ with ndf = $\nu_0 - \nu_1$ provided:
a) $H_0$ is true
b) $H_0$ and $H_1$ are nested
c) Params for $H_1 \rightarrow H_0$ are well defined, and not on boundary
d) Data is asymptotic
Wilks’ Theorem, contd

Examples: Does Wilks’ Th apply?

1) H0 = polynomial of degree 3
   H1 = polynomial of degree 5
   **YES: ΔS distributed as \( \chi^2 \) with \( \text{ndf} = (d-4) - (d-6) = 2 \)**

2) H0 = background only
   H1 = bgd + peak with free \( M_0 \) and cross-section
   **NO: H0 and H1 nested, but \( M_0 \) undefined when \( H1 \rightarrow H0 \). \( \Delta S \neq \chi^2 \)** (but not too serious for fixed M)

3) H0 = normal neutrino hierarchy
   H1 = inverted hierarchy
   **NO: Not nested. \( \Delta S \neq \chi^2 \) (e.g. can have \( \Delta \chi^2 \) negative)**

N.B. 1: Even when W. Th. does not apply, it does not mean that ΔS is irrelevant, but you cannot use W. Th. for its expected distribution.

N.B. 2: For large ndf, better to use ΔS, rather than \( S_1 \) and \( S_0 \) separately
Is difference in $S$ distributed as $\chi^2$?

Demortier:
H0 = quadratic bgd
H1 = ……………… +
Gaussian of fixed width, variable location & ampl

Protassov, van Dyk, Connors, ….
H0 = continuum
(a) H1 = narrow emission line
(b) H1 = wider emission line
(c) H1 = absorption line

Nominal significance level = 5%
Is difference in $S$ distributed as $\chi^2$ ?, contd.

So need to determine the $\Delta S$ distribution by Monte Carlo

N.B.

1) For mass spectrum, determining $\Delta S$ for hypothesis $H_1$ when data is generated according to $H_0$ is not trivial, because there will be lots of local minima

2) If we are interested in $5\sigma$ significance level, needs lots of MC simulations (or intelligent MC generation)

Background systematics

CMS Preliminary
\(\sqrt{s} = 7\) TeV, L = 5.1 fb\(^{-1}\)
\(\sqrt{s} = 8\) TeV, L = 5.3 fb\(^{-1}\)

- S/B Weighted Data
- S+B Fit
- Bkg Fit Component
- \(\pm 1\) \(\sigma\)
- \(\pm 2\) \(\sigma\)
Background systematics, contd

Signif from comparing $\chi^2$’s for H0 (bgd only) and for H1 (bgd + signal)
Typically, bgd = functional form $f_a$ with free params
   e.g. 4th order polynomial
Uncertainties in params included in signif calculation
   But what if functional form is different? e.g. $f_b$

Typical approach:
   If $f_b$ best fit is bad, not relevant for systematics
   If $f_b$ best fit is $\sim$comparable to $f_a$ fit, include contribution to systematics
   But what is ‘$\sim$comparable’?

Other approaches:
   Profile likelihood over different bgd parametric forms

Background subtraction
sPlots
Non-parametric background
Bayes
   etc

No common consensus yet among experiments on best approach
{Spectra with multiple peaks are more difficult}
“Handling uncertainties in background shapes: the discrete profiling method”

Dauncey, Kenzie, Wardle and Davies (Imperial College, CMS)


Has been used in CMS analysis of $H \rightarrow \gamma \gamma$

Problem with ‘Typical approach’: Alternative functional forms do or don’t contribute to systematics by hard cut, so systematics can change discontinuously wrt $\Delta \chi^2$

Method is like profile $\mathcal{L}$ for continuous nuisance params

Here ‘profile’ over discrete functional forms
Contours of $\ln L(s, \nu)$

$s = \text{physics param}$

$\nu = \text{nuisance param}$

Stat uncertainty on $s$ from width of $L$ fixed at $\nu_{\text{best}}$

Total uncertainty on $s$ from width of $L(s, \nu_{\text{prof}(s)}) = L_{\text{prof}}$

$\nu_{\text{prof}(s)}$ is best value of $\nu$ at that $s$

$\nu_{\text{prof}(s)}$ as fn of $s$ lies on green line

Total uncert $\geq$ stat uncertainty
$-2 \ln \mathcal{L}$

$\Delta$
Red curve: Best value of nuisance param $\nu$

Blue curves: Other values of $\nu$

Horizontal line: Intersection with red curve $\Rightarrow$ statistical uncertainty

‘Typical approach’: Decide which blue curves have small enough $\Delta$
Systematic is largest change in minima wrt red curves’.

Profile L: Envelope of lots of blue curves
Wider than red curve, because of systematics ($\nu$)
For $\mathcal{L} =$ multi-D Gaussian, agrees with ‘Typical approach’

Dauncey et al use envelope of finite number of functional forms
Point of controversy!
Two types of ‘other functions’:
a) Different function types e.g.
\[ \sum a_i x_i \] versus \[ \sum a_i / x_i \]
b) Given fn form but different number of terms
DDKW deal with b) by \(-2\ln L \rightarrow -2\ln L + kn\)
\[ n = \text{number of extra free params wrt best} \]
\[ k = 1, \text{ as in AIC (= Akaike Information Criterion)} \]

Opposition claim choice \( k=1 \) is arbitrary.
DDKW agree but have studied different values, and say \( k = 1 \) is optimal for them.
Also, any parametric method needs to make such a choice
p_0 \text{ v } p_1 \text{ plots}

Preprint by Luc Demortier and LL, “Testing Hypotheses in Particle Physics: Plots of p_0 versus p_1”
http://arxiv.org/abs/1408.6123

For hypotheses $H_0$ and $H_1$, $p_0$ and $p_1$ are the tail probabilities for data statistic $t$

Provide insights on:
- CLs for exclusion
- Punzi definition of sensitivity
- Relation of $p$-values and Likelihoods
- Probability of misleading evidence
- Sampling to foregone conclusion
- Jeffreys-Lindley paradox
CLs = \( \frac{p_1}{1-p_0} \) \rightarrow \text{diagonal line}

Provides protection against excluding \( H_1 \) when little or no sensitivity

Punzi definition of sensitivity:
Enough separation of pdf’s for no chance of ambiguity

\[ \Delta \mu \]

Can read off power of test
\text{e.g. If } H_0 \text{ is true, what is prob of rejecting } H_1? \]

\text{N.B. } p_0 = \text{tail towards } H_1
p_1 = \text{tail towards } H_0
Why $p \neq \text{Likelihood ratio}$

Measure different things:
$p_0$ refers just to $H_0$; $L_{01}$ compares $H_0$ and $H_1$

 Depends on amount of data:
e.g. Poisson counting expt little data:  
For $H_0$, $\mu_0 = 1.0$. For $H_1$, $\mu_1 = 10.0$
Observe $n = 10$ $p_0 \sim 10^{-7}$ $L_{01} \sim 10^{-5}$

Now with 100 times as much data, $\mu_0 = 100.0$ $\mu_1 = 1000.0$
Observe $n = 160$ $p_0 \sim 10^{-7}$ $L_{01} \sim 10^{+14}$
Jeffreys-Lindley Paradox

\( H_0 = \text{simple}, \quad H_1 \text{ has } \mu \text{ free} \)

\( p_0 \) can favour \( H_1 \), while \( B_{01} \) can favour \( H_0 \)

\[ B_{01} = \frac{L_0}{\int L_1(s) \pi(s) \, ds} \]

Likelihood ratio depends on signal:
e.g. Poisson counting expt small signal \( s \):
- For \( H_0 \), \( \mu_0 = 1.0 \).
- For \( H_1 \), \( \mu_1 = 10.0 \)
- Observe \( n = 10 \) \( p_0 \sim 10^{-7} \) \( L_{01} \sim 10^{-5} \) and favours \( H_1 \)

Now with 100 times as much signal \( s \), \( \mu_0 = 100.0 \) \( \mu_1 = 1000.0 \)
- Observe \( n = 160 \) \( p_0 \sim 10^{-7} \) \( L_{01} \sim 10^{+14} \) and favours \( H_0 \)

\( B_{01} \) involves integration over \( s \) in denominator, so a wide enough range
will result in favouring \( H_0 \)

However, for \( B_{01} \) to favour \( H_0 \) when \( p_0 \) is equivalent to \( 5\sigma \), integration
range for \( s \) has to be \( O(10^6) \) times Gaussian widths
WHY LIMITS?

Michelson-Morley experiment → death of aether

HEP experiments: If UL on expected rate for new particle < expected, exclude particle

CERN CLW (Jan 2000)
FNAL CLW (March 2000)
Heinrich, PHYSTAT-LHC, “Review of Banff Challenge”
Methods (no systematics)

Bayes (needs priors e.g. \( \text{const}, 1/\mu, \ 1/\sqrt{\mu}, \ \mu, \ldots \))

Frequentist (needs ordering rule, possible empty intervals, F-C)

Likelihood (DON’T integrate your L)

\( \chi^2 (\sigma^2 = \mu) \)

\( \chi^2 (\sigma^2 = n) \)

Recommendation 7 from CERN CLW: “Show your L”

1) Not always practical
2) Not sufficient for frequentist methods
DESIRABLE PROPERTIES

• Coverage
• Interval length
• Behaviour when $n < b$
• Limit increases as $\sigma_b$ increases
• Unified with discovery and interval estimation
90% Classical interval for Gaussian

$\sigma = 1 \quad \mu \geq 0 \quad \text{e.g. } m^2(\nu_e)$

$X_{\text{obs}} = 3$   Two-sided range

$X_{\text{obs}} = 1$   Upper limit

$X_{\text{obs}} = -2$  No region for $\mu$
Wants to avoid empty classical intervals

Uses “$L$-ratio ordering principle” to resolve ambiguity about “which 90% region?”

[Neyman + Pearson say $L$-ratio is best for hypothesis testing]

Unified  $\Rightarrow$ No ‘Flip-Flop’ problem
Classical (Neyman) Confidence Intervals

Uses only $P(\text{data}|\text{theory})$

Example:
- **Param** = Temp at centre of Sun
- **Data** = Est. flux of solar neutrinos

Theoretical Parameter $\mu$

$\mu \geq 0$

No prior for $\mu$

$\text{Prob}(\mu_l < \mu < \mu_u) = \alpha$
Classical (Neyman) Confidence Intervals

Uses only $P(\text{data}|\text{theory})$

Example:

Param = Temp at centre of Sun
Data = est. flux of solar neutrinos

$\mu \geq 0$

No prior for $\mu$
\( X_{\text{obs}} = -2 \) now gives upper limit
Search for Higgs:
H→γγ: low S/B, high statistics
$H \rightarrow Z\ Z \rightarrow 4\ l$: high S/B, low statistics
p-value for ‘No Higgs’ versus $m_H$
Mass of Higgs: Likelihood versus mass

-2 \Delta \ln L

m_\chi (GeV)

H \rightarrow \gamma\gamma + H \rightarrow ZZ

Combined
H \rightarrow \gamma\gamma
H \rightarrow ZZ

CMS Preliminary \( \sqrt{s} = 7 \text{ TeV}, L \leq 5.1 \text{ fb}^{-1} \) \( \sqrt{s} = 8 \text{ TeV}, L \leq 12.2 \text{ fb}^{-1} \)

-2 \Delta \ln L

m_\chi (GeV)

H \rightarrow \gamma\gamma + H \rightarrow ZZ

with syst.
no syst.
Comparing $0^+$ versus $0^-$ for Higgs

(like Neutrino Mass Hierarchy)

Conclusions

**Resources:**

Software exists: e.g. RooStats

Books exist: Barlow, Cowan, James, Lyons, Roe,.....


PDG sections on Prob, Statistics, Monte Carlo

CMS and ATLAS have Statistics Committees (and BaBar and CDF earlier) – see their websites

Before re-inventing the wheel, try to see if Statisticians have already found a solution to your statistics analysis problem.

Don’t use a square wheel if a circular one already exists.

"Good luck"