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 ln(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 and Lorenzo Moneta
Imperial College, London & Oxford
CERN
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:
p-values
  What they mean
  Combining p-values
Significance
Blind Analysis
LEE = Look Elsewhere Effect
Why 5σ for discovery?
Wilks’ Theorem
Background Systematics
\( p_0 \) v \( p_1 \) plots
Higgs search: Discovery, mass and spin

Conclusions
Examples of Hypotheses

1) Event selector  
(Event = particle interaction)

Events produced at CERN LHC at enormous rate
Online ‘trigger’ to select events for recording (~1 kiloHertz)
   e.g. events with many particles
Offline selection based on required features
   e.g. H0: Event contains top  H1: No top
Possible outcomes:  Events assigned as H0 or H1

2) Result of experiment

   e.g. H0 = nothing new
   H1 = new particle produced as well
      (Higgs, SUSY, 4\textsuperscript{th} neutrino,.....)

Possible outcomes  
\begin{tabular}{c|c|c}
   & H0 & H1 \\
   \hline
   ✓ & ✓ & X \\
   ✓ & X & ✓ \\
   ✓ & ✓ & ✓ \\
   X & X & ? \\
\end{tabular}

WRONG DECISIONS
E1:  Reject H0 when H0 true  (Loss of effic in 1))
E2:  Fail to reject H0 when H1 true  (Contamination)
H0 or H0 versus H1?

H0 = null hypothesis
e.g. Standard Model, with nothing new
H1 = specific New Physics     e.g. Higgs with $M_H = 125$ GeV
H0: “Goodness of Fit” e.g. $\chi^2$, p-values
H0 v H1: “Hypothesis Testing” e.g. $\mathcal{L}$-ratio
Measures how much data favours one hypothesis wrt other

H0 v H1 likely to be more sensitive for H1
Choosing between 2 hypotheses

Possible methods:

\[ \Delta \chi^2 \]
- p-value of statistic

\[ \ln L \text{–ratio} \]

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.
p-values

Concept of pdf

Example: Gaussian

\[ y = \frac{1}{\sqrt{(2\pi)\sigma}} \exp\left\{ -\frac{0.5(x-\mu)^2}{\sigma^2} \right\} \]

p-value: probability that \( x \geq x_0 \)

Gives probability of “extreme” values of data (in interesting direction)

<table>
<thead>
<tr>
<th>((x_0-\mu)/\sigma)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p)</td>
<td>16%</td>
<td>2.3%</td>
<td>0.13%</td>
<td>0.003%</td>
<td>0.3*10^{-6}</td>
</tr>
</tbody>
</table>

i.e. Small \(p\) = unexpected
p-values, contd

Assumes:
  Specific pdf for $x$ (e.g. Gaussian, no long tails)
  Data is unbiased
  $\sigma$ is correct

If so, and $x$ is from that pdf $\rightarrow$ uniform p-distribution

(Events at large $x$ give small $p$)
p-values for non-Gaussian distributions

e.g. Poisson counting experiment, bgd = b

\[ P(n) = e^{-b} \times b^n/n! \]

\{P = probability, not prob density\}

For \( n=7, p = \text{Prob( at least 7 events)} = P(7) + P(8) + P(9) + \ldots \ldots = 0.03 \]
Significance

Significance $= \frac{S}{\sqrt{B}}$ 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 $S = 0.1, B = 10^{-4}, \frac{S}{\sqrt{B}} = 10$

CONCLUSION:

Calculate $p$ properly (and allow for LEE if necessary)
p-values and $\sigma$

p-values often converted into equivalent Gaussian $\sigma$
e.g. $3 \times 10^{-7}$ is “5$\sigma$” (one-sided Gaussian tail)Does NOT imply that pdf = Gaussian
(Simply easier to remember number of $\sigma$, than p-value.)
What p-values are (and are not)

Reject H0 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 \( \text{prob(Theory;data)} \) related to \( \text{prob(data;Theory)} = \text{Likelihood} \)

by Bayes Th, including Bayesian prior

\[ P(\text{A;B}) \neq P(\text{B;A}) \]

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’]
Criticisms of p-values

(p-values banned by journal *Basic and Applied Social Psychology*)

1) Misunderstood

   So ban relativity, matrices.....?

2) Incorrect statements

3) p-values smaller than $\mathcal{L}$-ratios

   Measure different quantities
   p is only for one hypothesis
   $\mathcal{L}$-ratio compares two hypotheses
   (Is length or mass ‘better’ for comparing mouse and elephant?)
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 \( p_1 \times p_2 \times p_3 \ldots \)

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

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

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

Problems:

1) Recipe is not unique (Uniform dist in n-D hypercube \( \rightarrow \) uniform in 1-D)
2) Formula is not associative
   Combining \( \{\{p_1 \text{ and } p_2\} \), and then \( p_3 \}) \text{ gives different answer} \)
   from \( \{\{p_3 \text{ and } p_2\} \), 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 ************
Procedure for choosing between 2 hypotheses

1) No sensitivity
   - $H_0$ vs $H_1$

2) Maybe
   - $t$ vs $t_{crit}$

3) Easy separation
   - $t$ vs $t_{crit}$

Procedure:
- Obtain expected distributions for data statistic (e.g. $L$-ratio) for $H_0$ and $H_1$
- Choose $\alpha$ (e.g. 95%, 3\sigma, 5\sigma) and CL for $p_1$ (e.g. 95%)
- Given $b$, $\alpha$ determines $t_{crit}$
- $b+s$ defines $\beta$. For $s > s_{min}$, separation of curves $\rightarrow$ discovery or excln
- $1-\beta$ = Power of test

Now data:
- If $t_{obs} \geq t_{crit}$ (i.e. $p_0 \leq \alpha$), discovery at level $\alpha$
- If $t_{obs} < t_{crit}$, no discovery. If $p_1 < 1-CL$, exclude $H_1$ (or CLs = $p_1/(1-p_0)$)

For event selector, $1-\alpha$ = efficiency for signal events; $\beta$ = mis-ID prob from other events
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

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.

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
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 that 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 5σ for Discovery?

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

Our reasons:

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

\[
\begin{align*}
p(H_1 | x) &= \frac{p(x | H_1) \times \pi(H_1)}{p(x | H_0) \times \pi(H_0)} \\
p(H_0 | x) &= \frac{p(x | H_0)}{\pi(H_0)}
\end{align*}
\]

Posterior Likelihood Priors
prob ratio

“Extraordinary claims require extraordinary evidence”

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

Alternative suggestion:

<table>
<thead>
<tr>
<th>SEARCH</th>
<th>SURPRISE</th>
<th>IMPACT</th>
<th>LEE</th>
<th>SYSTEMATICS</th>
<th>No. σ</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>Δm</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Neutrino osc</td>
<td>Medium</td>
<td>High</td>
<td>sin^22θ, Δm^2</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>B_s → μμ</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)_μ anom</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>H spin ≠ 0</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>Medium</td>
<td>5</td>
</tr>
<tr>
<td>4th gen q, l, ν</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 `carved in stone on Mt. Sinai’

Bob Cousins: “2 independent expts each with 3.5σ better than one expt with 5σ”
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) \( H_0 = \) polynomial of degree 3
   \( H_1 = \) polynomial of degree 5
   **YES: \( \Delta S \) distributed as \( \chi^2 \) with \( \text{ndf} = (d-4) - (d-6) = 2 \)**

2) \( H_0 = \) background only
   \( H_1 = \) bgd + peak with free \( M_0 \) and cross-section
   **NO: \( H_0 \) and \( H_1 \) nested, but \( M_0 \) undefined when \( H_1 \rightarrow H_0 \). \( \Delta S = \chi^2 \) (but not too serious for fixed \( M \))

3) \( H_0 = \) normal neutrino hierarchy
   \( H_1 = \) inverted hierarchy
   **NO: Not nested. \( \Delta S = \chi^2 \) (e.g. can have \( \Delta S \) negative)**

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

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

Demortier:
\(H_0 = \text{quadratic bgd}\)
\(H_1 = \text{Gaussian of fixed width, variable location} \& \text{ampl}\)

What is peak at zero?
Why not half the entries?

Protassov, van Dyk, Connors, ....
\(H_0 = \text{continuum}\)
(a) \(H_1 = \text{narrow emission line}\)
(b) \(H_1 = \text{wider emission line}\)
(c) \(H_1 = \text{absorption line}\)

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 H1 when data is generated according to H0 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

\( \sqrt{s} = 7 \text{ TeV}, L = 5.1 \text{ fb}^{-1} \)

\( \sqrt{s} = 8 \text{ TeV}, L = 5.3 \text{ fb}^{-1} \)

- S/B Weighted Data
- S+B Fit
- Bkg Fit Component
- ±1σ
- ±2σ
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 ~comparable to $f_a$ fit, include contribution to systematics
  But what is ‘~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→γγ

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

Method is like profile L for continuous nuisance params

Here ‘profile’ over discrete functional forms
Reminder of Profile $\mathcal{L}$

Contours of $\ln \mathcal{L}(s, \nu)$

$s = $ physics param
$\nu = $ nuisance param

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

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

$\nu_{\text{prof}(s)}$ 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. 
\[ \Sigma a_i \times_i \text{ versus } \Sigma a_i/x_i \]
b) Given fn form but different number of terms
DDKW deal with b) by \(-2\ln L \rightarrow -2\ln L + \text{kn}\)
\[ n = \text{number of extra free params wrt best} \]
\[ k = 1, \text{ as in AIC (= Akaike Information Criterion)} \]

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
$p_0 \, v \, p_1$ plots


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
- Jeffreys-Lindley paradox
\[ \text{CLs} = \frac{p_1}{1-p_0} \rightarrow \text{diagonal line} \]

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

Can read off power of test
e.g. If \( H_0 \) is true, what is prob of rejecting \( H_1 \)?

**N.B.** \( p_0 \) = tail towards \( H_1 \)
\( p_1 \) = tail towards \( H_0 \)
\( \alpha, \beta, \text{Errors of 1}^{\text{st}} \text{ and 2}^{\text{nd}} \text{ Kind, etc.} \)

e.g. \( H_0 = \text{event with top} \quad H_1 = \text{no top} \)

\( \alpha = \text{prob of rejecting } H_0 \text{ when } H_0 \text{ true} = E_1 \)
\[ p_0 < \alpha, \text{ reject as top event} \]
\[ p_0 > \alpha, \text{ accept as top event} \]
\( \text{Effic for } H_0 = 1 - \alpha \)

\( \beta = \text{value of } p_1 \text{ when } p_0 = \alpha \)
\[ = \text{prob of not rejecting } H_0 \text{ when } H_1 \text{ true} = E_2 \]
\[ = \text{mis-ID of ‘no top’ events} \]

\( \text{Power} = \text{prob of rejecting } H_0 \text{ when } H_1 \text{ true} = 1 - \beta \)
\( \text{Contamination in signal sample depends on } \beta, \text{ and relative frequencies for } H_0 \text{ and } H_1 \text{ events.} \)

\( \text{ROC curves plot ‘1- Bgd Mis-ID’ versus ‘Signal Efficiency’} \)
\[ = \text{‘1- } p_1 \text{’ versus ‘1- } p_0 \text{’} \quad (\text{Cf } p_1 \text{ v } p_0 \text{ plots}) \)
Why $p \neq$ Likelihood ratio

Measure different things:
$p_0$ refers just to H0; $L_{01}$ compares H0 and H1

Depends on amount of data:
e.g. Poisson counting expt little data:
   For H0, $\mu_0 = 1.0$. For H1, $\mu_1 = 10.0$
   Observe $n = 10$ \hspace{1cm} $p_0 \sim 10^{-7}$ \hspace{1cm} $L_{01} \sim 10^{-5}$

Now with 100 times as much data, $\mu_0 = 100.0$ \hspace{1cm} $\mu_1 = 1000.0$
   Observe $n = 160$ \hspace{1cm} $p_0 \sim 10^{-7}$ \hspace{1cm} $L_{01} \sim 10^{+14}$

N.B. In HEP, data statistic is typically $L_{01}$

Can think of method as:
p-value, where data statistic just happens to be $L_{01}$; or
$L_{01}$ method where p-values are just used for calibration.
Jeffreys-Lindley Paradox

H_0 = simple, H_1 has μ 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, μ_0 = 1.0. For H_1, μ_1 =10.0
   Observe n = 10 \quad p_0 \sim 10^{-7} \quad L_{01} \sim 10^{-5} \quad \text{and favours } H_1
Now with 100 times as much signal s, μ_0 = 100.0 \quad μ_1 =1000.0
   Observe n = 160 \quad p_0 \sim 10^{-7} \quad L_{01} \sim 10^{+14} \quad \text{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 5σ, integration
range for s has to be \(O(10^6)\) times Gaussian widths
WHY LIMITS?

Michelson-Morley experiment $\rightarrow$ death of aether

HEP experiments:
If UL on expected rate for new particle $< \text{expected}$, exclude particle
Do as function of $M_X \rightarrow$ excluded mass range below $M_e$

Compare with expected $M_e \rightarrow \text{expt’s sensitivity}$

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

Bayes (needs priors e.g. const, $1/\mu$, $1/\sqrt{\mu}$, $\mu$, .....)
Frequentist (needs ordering rule,
    possible empty intervals, F-C)

CLs
Likelihood (DON’T integrate your $\mathcal{L}$)
$\chi^2 (\sigma^2 = \mu)$
$\chi^2 (\sigma^2 = n)$

Recommendation 7 from CERN CLW: “Show your $\mathcal{L}$”
1) Not always practical
2) Not sufficient for frequentist methods
Ilya Narsky, FNAL CLW 2000

Poisson counting expt

\[ b = 3.0 \]
Search for Higgs: $H \rightarrow \gamma \gamma$: low S/B, high statistics
H→Z Z → 4 l: high S/B, low statistics
p-value for ‘No Higgs’ versus $m_H$
Mass of Higgs: Likelihood versus mass

- $\Delta -2 \ln L$

$m_X$ (GeV)

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

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

- Combined
- $H \rightarrow \gamma \gamma$
- $H \rightarrow ZZ$

- $\Delta -2 \ln L$

$m_X$ (GeV)
Comparing $0^+$ versus $0^-$ for Higgs (like Neutrino Mass Hierarchy)

Conclusions

**Resources:**

Software exists: e.g. RooStats

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

New: `Data Analysis in HEP: A Practical Guide to Statistical Methods`, Behnke et al.

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”