This presentation would have not been possible without the tremendous help of the following people throughout many years:

- Louis Lyons
- Alex Read
- Glen Cowan
- Kyle Cranmer
- Ofer Vitells
- Bob Cousins
Application of CIs and $q^{NP}$ test statistic

$$q^{NP} = -2 \ln \frac{L(H_0)}{L(H_1)} = \sum_{\text{bins}} -2 \ln \frac{L_i(0^+)}{L_i(0^-)}$$

$$L_i(0^+) = \text{Pois}(n_i; n_i^{0^+}) = \left( \frac{n_i^{0^+}}{n_i} \right)^{n_i} e^{-n_i^{0^+}}$$

Can you tell $O^+$ from $0^-$?

ATLAS

$H \rightarrow ZZ^* \rightarrow 4l$

- $J^p = 0^+$
- $J^p = 0^-$

$\sqrt{s} = 7 \text{ TeV}$
$\int L dt = 4.6 \text{ fb}^{-1}$

$\sqrt{s} = 8 \text{ TeV}$
$\int L dt = 20.7 \text{ fb}^{-1}$
Test Spin 0 parity

\[ H_0 = 0^+ \]

\[ H_1 = 0^- \]

\[ p_{H_1}(\text{exp} \mid H_0) = 0.37\% , \]
\[ p_{H_1}(\text{obs}) = 1.5\% \]
\[ p_{H_0}(\text{obs}) = 31\% \]
\[ p_{H_1}^{CL_s}(\text{obs}) = 2.2\% \]

\[ p_{H_1}^{CL_s} = \frac{p_{H_1}}{1 - p_{H_0}} = \frac{1.5\%}{1 - 0.31} = 2.2\% \]

Which means \( J^p=0^- \) is excluded at the 97.8% CL in favour of \( J^p=0^+ \)

\[ q^{NP} = -2 \ln \frac{L(H_0)}{L(H_1)} \]

\[ Eilam Gross ESHEP \]

September 2017
Case Study: Higgs Discovery
Basic Definition: Signal Strength

- We normally relate the signal strength to its expected Standard Model value, i.e.

\[ \mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)} \]

\[ \hat{\mu}(m_H) = \text{MLE of } \mu \]
Introducing the Heartbeat

\[ \mu(m_H) = \frac{\sigma(m_H)}{\sigma_{SM}(m_H)} \]

\[ \hat{\mu}(m_H) = \text{MLE of } \mu \]
Reminder: The test statistic

\[ q_0 = \begin{cases} 
-2 \ln \lambda(0) & \hat{\mu} \geq 0, \\
0 & \hat{\mu} < 0,
\end{cases} \]

\[ q_{\mu} = \begin{cases} 
-2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\
0 & \hat{\mu} > \mu
\end{cases} \]

- Downward fluctuations of the background do not serve as an evidence against the background
- Upward fluctuations of the signal do not serve as an evidence against the signal
p₀ and the expected p₀

\[ p₀ = \int_{q_{0,\text{obs}}}^{\infty} f(q₀ | 0) \, dq₀ \]

p₀ is the probability to observe a less BG like result (more signal like) than the observed one.

Small p₀ leads to an observation.
A tiny p₀ leads to a discovery.

\[ p = \int_{Z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx = 1 - \Phi(Z) \]

\[ Z = \Phi^{-1}(1 - p) \]
Distribution of q0 (discovery)

- We find

\[ f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2\sqrt{2\pi}} \frac{1}{\sqrt{q_0}} e^{-q_0/2}. \]

\[ p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0) \, dq_0. \]

\[ Z_0 = \Phi^{-1}(1-p_0) = \sqrt{q_0}. \]

- \( q_0 \) distribute as half a delta function at zero and half a chi squared. \( q_{0,\text{obs}} = q_{0,\text{obs}}(m_H) \rightarrow p_0 = p_0(m_H) \)
The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

$$Z_A = \sqrt{q_{0,A}}$$

$$\text{med}[Z_0|1] = \sqrt{q_{0,A}} = \sqrt{2 ((s + b) \ln(1 + s/b) - s)}$$

$$Z_A = \sqrt{q_{0,A}} \xrightarrow{s/b \ll 1} \frac{s}{\sqrt{b}} + O(s/b)$$

Eilam Gross ESHEP

September 2017
The New $s/\sqrt{b}$

$s/\sqrt{b}$?

The new $s/\sqrt{b}$

$$\text{med}[Z_0|1] = \sqrt{q_{0,A}} = \sqrt{2\left((s + b) \ln\left(1 + s/b\right) - s\right)}$$
Taking Background Systematics into Account

- The intuitive explanation of $s/\sqrt{b}$ is that it compares the signal, $s$, to the standard deviation of $n$ assuming no signal, $\sqrt{b}$.

- Now suppose the value of $b$ is uncertain, characterized by a standard deviation $\sigma_b$.

- A reasonable guess is to replace $\sqrt{b}$ by the quadratic sum of $\sqrt{b}$ and $\sigma_b$, i.e.,

$$b \pm \Delta \cdot b \Rightarrow \sigma_b = \sqrt{\left(\sqrt{b}\right)^2 + \left(\Delta \cdot b\right)^2} = \sqrt{b + \Delta^2 b^2}$$

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

$$\frac{s}{b} \geq 5 \Rightarrow \frac{s}{b} \geq 0.5 \text{ for } \Delta \sim 10\%$$

If $s/b < 0.5$ we will never be able to make a discovery

But even that formula can be improved using the Asimov formalism
Significance with systematics

- We find (G. Cowan)

\[
Z_A = \left[ 2 \left( (s + b) \ln \left( \frac{(s + b)(b + \sigma_b^2)}{b^2 + (s + b)\sigma_b^2} \right) - \frac{b^2}{\sigma_b^2} \ln \left( 1 + \frac{\sigma_b^2 s}{b(b + \sigma_b^2)} \right) \right) \right]^{1/2}
\]

Expanding the Asimov formula in powers of \( s/b \) and \( \sigma_b^2/b \) gives

\[
Z_A = \frac{s}{\sqrt{b + \sigma_b^2}} \left( 1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)
\]

- So the “intuitive” formula can be justified as a limiting case of the significance from the profile likelihood ratio test evaluated with the Asimov data set.
Significance with systematics

\begin{align*}
  s &= 5 \\
  \sigma_b / b &= 0.2, 0.5 \\
  \frac{s}{\sqrt{b + \sigma_b^2}} &\quad \text{Z}_A \\
  \text{Monte Carlo}
\end{align*}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{chart.png}
\end{figure}
Example: $H \rightarrow \gamma \gamma$

Selected diphoton sample
- Data 2011 and 2012
- $\text{Sig} + \text{Bkg}$ inclusive fit ($m_H = 126.5 \text{ GeV}$)
- 4th order polynomial

$\sqrt{s} = 7 \text{ TeV}, \int L dt = 4.8 \text{ fb}^{-1}$
$\sqrt{s} = 8 \text{ TeV}, \int L dt = 5.9 \text{ fb}^{-1}$

ATLAS Preliminary
$H \rightarrow \gamma\gamma$

**ATLAS** Preliminary

Data 2011, $\sqrt{s} = 7$ TeV, $\int L dt = 4.8$ fb$^{-1}$

Data 2012, $\sqrt{s} = 8$ TeV, $\int L dt = 5.9$ fb$^{-1}$
$H \rightarrow \gamma \gamma$

**ATLAS** Preliminary

- Data 2011, $\sqrt{s} = 7$ TeV, $\int L dt = 4.8$ fb$^{-1}$
- Data 2012, $\sqrt{s} = 8$ TeV, $\int L dt = 5.9$ fb$^{-1}$

$\text{SM } H \rightarrow \gamma \gamma$
Obtaining the Syst Error

\[ \sigma_{syst} = \sqrt{\sigma_{tot}^2 - \sigma_{stat}^2} \]
Measurements

Case studies: ATLAS and CMS
mass and coupling combinations
PL in obtaining the mass

\[ \Lambda(\alpha) = \frac{L(\alpha, \hat{\theta}(\alpha))}{L(\hat{\alpha}, \hat{\theta})} \quad t_\alpha = -2\ln \Lambda(\alpha) \]

\[ \Lambda(m_H) = \frac{L(m_H, \hat{\mu}_{ggF+t\bar{t}H(\bar{b}\bar{b}H)}(m_H), \hat{\mu}_{VBF+V_H}(m_H), \hat{\mu}_{ZZ}(m_H), \hat{\theta}(m_H))}{L(\hat{m}_H, \hat{\mu}_{ggF+t\bar{t}H(\bar{b}\bar{b}H)}, \hat{\mu}_{VBF+V_H}, \hat{\mu}_{ZZ}, \hat{\theta})} \]

Scan the test statistic \( t_\alpha = t(\alpha) \)

find \( \hat{\alpha} \)

\[ t(\hat{\alpha} \pm N\sigma_{\hat{\alpha}}) = N^2 \]
A case of 2 poi

• In order to address the values of the signal strength and mass of a potential signal that are simultaneously consistent with the data, the following profile likelihood ratio is used:

\[
\lambda(\mu, m_H) = \frac{L(\mu, m_H, \hat{\theta}(\mu, m_H))}{L(\hat{\mu}, \hat{m}_H, \hat{\theta})}
\]

• In the presence of a signal, this test statistic will produce closed contours about the best fit point \((\hat{\mu}, \hat{m}_H)\);

• The 2D LR behaves asymptotically as a Chis squared with 2 DOF (Wilks’ theorem) so the derivation of 68% and 95% CL cintours is easy, but care must be taken; The projection of 2D CI are not 1D CI!
PL in obtaining the Couplings

\[ \Lambda(k_F, k_V) = \frac{L(\hat{k}_F, k_V, \hat{\theta}(k_F, k_V))}{L(\hat{k}_F, \hat{k}_V, \hat{\theta})}. \]
68% CI is a tricky issue

Is the WW a better measurement than the combination?

1D CI
Is not
2D CI

ATLAS and CMS
Run 1 Internal

-2 ln \( \lambda (k_F^{WW}) \)

-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

\( k_F^{WW} \)

(c)
68% CI is a tricky issue

Is the WW a better measurement than the combination?
1D vs 2D Confidence Interval

\[ \Delta \chi^2 = 1 \]
\[ \Delta \chi^2 = 2.3 \ (68\% \ CL) \]
Multidimensional PL
A Tutorial
A toy case with 1-3 poi

3 cases studied
1poi: $\mu$ while $\epsilon, A, b$ profiled
2poi: $\mu, \epsilon$ profile $A$ and $b$
3poi: $\mu, \epsilon, A$ profile $b$

\[
n = \mu \epsilon A_s + b
\]
\[
L = L(\mu, \epsilon, A, b)
\]
\[
L(\mu, \epsilon, A) = \frac{(\mu \epsilon A_s + b)^n}{n!} e^{- (\mu \epsilon A_s + b)} \frac{1}{\sigma_\epsilon \sqrt{2\pi}} e^{- (\epsilon_{\text{meas}} - \epsilon)^2 / 2\sigma_\epsilon^2} \frac{1}{\sigma_b \sqrt{2\pi}} e^{- (b_{\text{meas}} - b)^2 / 2\sigma_b^2} \frac{1}{\sigma_A \sqrt{2\pi}} e^{- (A_{\text{meas}} - A)^2 / 2\sigma_A^2}
\]
A toy case with 3 poi

$$L(\mu, \varepsilon, A) = \frac{\mu \varepsilon As + b)^n}{n!} e^{-\mu \varepsilon As + b) \frac{1}{\sigma_{\varepsilon}\sqrt{2\pi}} e^{-\left(\varepsilon_{meas} - \varepsilon\right)^2 / 2\sigma_{\varepsilon}^2} \frac{1}{\sigma_b\sqrt{2\pi}} e^{-\left(b_{meas} - b\right)^2 / 2\sigma_b^2} \frac{1}{\sigma_A\sqrt{2\pi}} e^{-\left(A_{meas} - A\right)^2 / 2\sigma_A^2}}$$

three parameters of interest (profiling only b) non-profiled parameters set to their real value

$$f(q(1)|\mu=1)$$

$$\chi^2_{3}$$

background = 100
signal = 90
$$\varepsilon = 0.5$$
$$A = 0.7$$
$$\sigma_{\varepsilon} = 0.05$$
$$\sigma_b = 10$$
$$\sigma_A = 0.2$$

6000 events

$$\chi^2(n_{dof}=3)$$  $$\chi^2(n_{dof}=2)$$  $$\chi^2(n_{dof}=1)$$
A toy case with 2 poi

\[ L(\mu, \varepsilon, A) = \frac{(\mu \varepsilon A \varepsilon + b)^n}{n!} e^{-(\mu \varepsilon A \varepsilon + b)} \frac{1}{\sigma \varepsilon \sqrt{2\pi}} e^{-(\varepsilon_{\text{meas}} - \varepsilon)^2 / 2\sigma^2} \frac{1}{\sigma b \sqrt{2\pi}} e^{-(b_{\text{meas}} - b)^2 / 2\sigma_b^2} \frac{1}{\sigma A \sqrt{2\pi}} e^{-(A_{\text{meas}} - A)^2 / 2\sigma_A^2} \]

Two parameters of interest (profiling A and b) non-profiled parameters set to their real value

\[ f(q(1) \mid \mu = 1) \]

\[ \chi^2 \]

\[ \chi^2(n_{\text{dof}} = 3) \]

\[ \chi^2(n_{\text{dof}} = 2) \]

\[ \chi^2(n_{\text{dof}} = 1) \]

Background = 100
Signal = 90
\[ \varepsilon = 0.5 \]
\[ A = 0.7 \]
\[ = 0.05 \]
\[ = 10 \]
\[ = 0.2 \]

6000 events

September 2017
A toy case with 1 poi

$$L(\mu, \varepsilon, A) = \frac{(\mu \varepsilon A_s + b)^n}{n!} e^{-(\mu \varepsilon A_s + b)} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-(\varepsilon_{\text{meas}} - \varepsilon)^2 / 2\sigma_\varepsilon^2} \frac{1}{\sigma_b \sqrt{2\pi}} e^{-(b_{\text{meas}} - b)^2 / 2\sigma_b^2} \frac{1}{\sigma_A \sqrt{2\pi}} e^{-(A_{\text{meas}} - A)^2 / 2\sigma_A^2}$$

The one parameter of interest (profiling $\varepsilon$ $A$ and $b$) non-profiled parameters set to their real value

- background = 100
- signal = 90
- $\varepsilon = 0.5$
- $A = 0.7$
- $\sigma_\varepsilon = 0.05$
- $\sigma_b = 10$
- $\sigma_A = 0.2$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

6000 events
Significance

For the fixed data set
The Nuisance Parameters
Are fixed to their nominal values.
The likelihood are more parabolic,
yet, never symmetric
The asymptotic hold!

\[ Z_{obs} = \sqrt{q_{0,obs}} \]

Asimov Data Set

\[ Z_A = \sqrt{4.5} \]

Random Data Set

\[ b_{meas} = 106.84 \]
\[ \varepsilon_{meas} = 0.523 \]
\[ A_{meas} = 0.477 \]
\[ n_{meas} = 121 \]

background = 100
signal = 90
\[ \varepsilon = 0.5 \]
\[ A = 0.7 \]
\[ \sigma_{\varepsilon} = 0.05 \]
\[ \sigma_b = 10 \]
\[ \sigma_A = 0.2 \]
Pulls and Ranking of NPs

The pull of $\theta_i$ is given by $\frac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}$

without constraint $\sigma\left(\frac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}\right) = 1$ $\left<\frac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}\right> = 0$

It’s a good habit to look at the pulls of the NPs and make sure that Nothing irregular is seen

In particular one would like to guarantee that the fits do not over constrain a NP in a non sensible way
Asimov

\[ s = 90 \]

- \[ b_{\text{meas}} = 100 \]
- \[ \varepsilon_{\text{meas}} = 0.5 \]
- \[ A_{\text{meas}} = 0.7 \]
- \[ \mu_{\text{meas}} = 1 \]
- \[ n_{\text{meas}} = \mu \varepsilon A + b = 131.5 \]

reminder:
- \[ b_0 = 100 \]
- \[ \varepsilon_0 = 0.5 \]
- \[ A_0 = 0.7 \]
- \[ \mu_0 = 1 \]
- \[ n_0 = 131.5 \]
- signal = 90

\[ \sigma_0 \]
- \[ \sigma_{\varepsilon} = 0.05 \]
- \[ \sigma_b = 10 \]
- \[ \sigma_A = 0.2 \]

To get the pulls:
- scan \( q(\varepsilon) \)
- Find \( \hat{\varepsilon} \)
- Find \( \sigma_+ \) and \( \sigma_- \) i.e. the positive and negative error bar substituting \( q(\varepsilon) = 1 \)

With the Asimov data sets we find perfect pulls for the profiled scans
But not for the fix scans!
Random Data Set

\[ n_{\text{meas}} = 132 \]
\[ b_{\text{meas}} = 103.208 \]
\[ \varepsilon_{\text{meas}} = 0.465459 \]
\[ A_{\text{meas}} = 0.487107 \]
\[ \mu_{\text{meas}} = 1.41099 \]

reminder:
\[ b_0 = 100 \]
\[ \varepsilon_0 = 0.5 \]
\[ A_0 = 0.7 \]
\[ \mu_0 = 1 \]
\[ n_0 = 131.5 \]
\[ \text{signal} = 90 \]

\[ \sigma_0 \]
\[ \sigma_\varepsilon = 0.05 \]
\[ \sigma_b = 10 \]
\[ \sigma_A = 0.2 \]

To get the pulls:

- scan \( q(\varepsilon) \)
- Find \( \hat{\varepsilon} \)
- Find \( \sigma^+ \) and \( \sigma^- \) i.e. the positive and negative error bar substituting \( q(\varepsilon) = 1 \)

With the random data sets we find perfect pulls for the profiled scans
But not for the fix scans!
To get the impact of a Nuisance Parameter in order to rank them:

Say we want the impact of $\epsilon$

- Scan $q(\epsilon)$, profiling all other NPs
- Find $\hat{\epsilon}$
- (note that $\hat{\mu}_{\hat{\epsilon}} = \hat{\mu}$)
- Find $\hat{\mu}_{\hat{\epsilon} \pm \sigma_{\hat{\epsilon}}} = \hat{\mu}_{\hat{\epsilon} \pm \sigma_{\hat{\epsilon}}}$
- The impact is given by $\Delta \mu^{\pm} = \hat{\mu}_{\hat{\epsilon} \pm \sigma_{\hat{\epsilon}}} - \hat{\mu}$
Asimov: SUMMARY of Pulls and Impact

\[ b_{\text{meas}} = 100 \]
\[ \epsilon_{\text{meas}} = 0.5 \]
\[ A_{\text{meas}} = 0.7 \]
\[ \mu_{\text{meas}} = 1 \]
\[ n_{\text{meas}} = \mu \epsilon A + b = 131.5 \]

Reminder:
\[ b_0 = 100 \]
\[ \epsilon_0 = 0.5 \]
\[ A_0 = 0.7 \]
\[ \mu_0 = 1 \]
\[ n_0 = 131.5 \]
\[ \text{signal} = 90 \]

\[ \sigma_0 \]
\[ \sigma_{\epsilon} = 0.05 \]
\[ \sigma_b = 10 \]
\[ \sigma_A = 0.2 \]
Random Data Set: SUMMARY of Pulls and Impact

- \( n_{\text{meas}} = 132 \)
- \( b_{\text{meas}} = 103.208 \)
- \( \varepsilon_{\text{meas}} = 0.465459 \)
- \( A_{\text{meas}} = 0.487107 \)
- \( \mu_{\text{meas}} = 1.41099 \)

Reminder:
- \( b_0 = 100 \)
- \( \varepsilon_0 = 0.5 \)
- \( A_0 = 0.7 \)
- \( \mu_0 = 1 \)
- \( n_0 = 131.5 \)
- \( \text{signal} = 90 \)

- \( \sigma_0 = 0.05 \)
- \( \sigma_{\varepsilon} = 0.05 \)
- \( \sigma_b = 10 \)
- \( \sigma_A = 0.2 \)

Diagram: 
- Yellow: negative correlation
- Light blue: positive correlation
Real Examples
Pulls and Ranking of NPs

Ranking $\theta_i$ by its effect in the NP

$$\Delta \mu^\pm = \hat{\mu} \pm \sigma^\pm - \hat{\mu}$$

By ranking we can tell which NPs are the important ones and which can be pruned.
The lack of events in spite of an expected background allows us to set a better limit than the expected.
E.G., O. Vitells “Trial factors for the look elsewhere effect in high energy physics”,

O. Vitells and E. G., Estimating the significance of a signal in a multi-dimensional search,

Search for resonances in diphoton events at $s\sqrt{=}13$ TeV with the ATLAS detector, ATLAS collaboration, arXiv 1606.03833
Look Elsewhere Effect

Is there a signal here?
Look Elsewhere Effect

Looks like @ m=30
What is its significance?
What is your test statistic?

\[ q_{\text{fix,obs}} = -2 \ln \frac{L(b)}{L(\hat{\mu}s(m=30) + b)} \]
Look Elsewhere Effect

Would you ignore this signal, had you seen it?
Look Elsewhere Effect

Or this?
Look Elsewhere Effect

Or this?
Look Elsewhere Effect

Or this?

Obviously NOT!

ALL THESE “SIGNS” ARE BG FLUCTUATIONS
Look Elsewhere Effect

Having no idea where the signal might be there are two options

**OPTION I:**
scan the mass range in pre-defined steps and test any disturbing fluctuations
(do not confuse me with the facts)
Perform a fixed mass analysis at each point

\[ q_{\text{fix,obs}}(\hat{\mu}) = -2 \ln \frac{L(b)}{L(\hat{\mu}s(m) + b)} \]
Look Elsewhere Effect

Having no idea where the signal might be there are two options

**OPTION I:**
scan the mass range in pre-defined steps and test any disturbing fluctuations
(do not confuse me with the facts)
Perform a fixed mass analysis at each point

\[
q_{\text{fix,obs}}(\hat{\mu}) = -2 \ln \frac{L(b)}{L(\hat{\mu}s(m) + b)}
\]
Look Elsewhere Effect

The scan resolution must be less than the signal mass resolution.

Assuming the signal can be only at one place, pick the one with the smallest p-value (maximum significance).

\[
q = \max_m \left\{ q_{\text{fix,obs}} (\hat{\mu}) \right\} = \max_m \left\{ -2 \ln \frac{L(b)}{L(\hat{\mu}_s(m) + b)} \right\}
\]

\[
= \min_m \left\{ p - \text{value} \right\}
\]
Look Elsewhere Effect

Test statistic

\[ q_{fix,obs} = -2 \ln \frac{L(b)}{L(\hat{\mu}s(m = 30) + b)} \]

What is the \( p \)-value?

generate the PDF

\[ f(q_{fix} \mid H_0) \]

and find the \( p \)-value

Wilks theorem:

\[ f(q_{fix} \mid H_0) \sim \chi^2 \]

\[ p_{fix} = \int_{q_{fix,obs}}^{\infty} f(q_{fix} \mid H_0) dq_{fix} \]
Look Elsewhere Effect: Floating Mass

Option II:
Leave the mass floating

Having no idea where the signal might be you would allow the signal to be anywhere in the search range and use a modified test statistic

For the same observation, the p-value increases because more possibilities are opened

\[ q_{\text{float,obs}}(\hat{\mu}, \hat{m}) = -2 \ln \frac{L(b)}{L(\hat{\mu}s(\hat{m}) + b)} \]
the test statistic

\[ q_{\text{float,obs}}(\hat{\mu}, \hat{m}) = -2 \ln \frac{L(b)}{L(\hat{\mu}s(\hat{m}) + b)} \]

The null hypothesis PDF

\[ f(q_{\text{float}} \mid H_0) \]

does not follow a chi-squared with 2 dof because there are multiple minima depending on the size of the search range and resolution.
Assume a maximal local fluctuation 
mass \( \hat{m} = 30 \)

We can calculate the following:

\[
q_{\text{fix,obs}} = q_{\text{float,obs}} = -2 \ln \frac{L(b)}{L(\hat{m} = m = 30) + b)}
\]

\[
p_{\text{fix}} = \int_{q_{\text{obs}}} f(q_{\text{fix}} | H_0) dq_{\text{fix}} < p_{\text{float}} = \int_{q_{\text{obs}}} f(q_{\text{float}} | H_0) dt_{\text{float}}
\]

\[
trial\# = \frac{\int_{q_{\text{obs}}} f(q_{\text{float}} | H_0) dt_{\text{float}}}{\int_{q_{\text{obs}}} f(q_{\text{fix}} | H_0) dt_{\text{fix}}} = \frac{p_{\text{float}}}{p_{\text{fix}}} > 1
\]

**Can we analytically calculate the trial\# (or p\_float)?**
Define the Problem

- Let \( n = \mu s(m, \Gamma) + b \)
- \( m, \Gamma \) are nuisance parameters undefined under the null hypothesis \( \mu = 0 \)
- What is the pdf of 
  \[
  \hat{q}_0 \equiv q_0(\hat{m}, \hat{\Gamma}) = -2 \log \frac{L(\mu = 0)}{L(\hat{\mu}, \hat{m}, \hat{\Gamma})} = \max[ q_0(m, \Gamma) ]
  \]
  under the null hypothesis
- To generalize the problem, let \( \Theta \) be the nuisance parameter, undefined under the null hypothesis, and let us try to find out the pdf of 
  \[
  \hat{q}_0 \equiv q_0(\hat{\Theta}) = -2 \log \frac{L(\mu = 0)}{L(\hat{\mu}, \hat{\Theta})} = \max[ q_0(\Theta) ]
  \]
  for which we want to calculate
  \[
  p\text{-value} = P(\max[ q_0(\Theta) ] \geq u), u = Z^2
  \]
The profile-likelihood test statistic
(with a nuisance parameter that is not defined under the Null hypothesis)

- Consider the test statistic:
  \[ q_0(\theta) = -2\log \frac{L(\mu = 0)}{L(\hat{\mu}, \theta)} \quad H_0 : \mu = 0 \]
  \[ H_1 : \mu > 0 \]

- For some fixed \( \theta \), \( q_0(\theta) \) has (asymptotically) a \( \chi^2 \) distribution with one degree of freedom by Wilks’ theorem.

- \( q_0(\theta) \) is a \textit{chi}^2 \textit{random field} over the space of \( \theta \) (a random variable indexed by a continuous parameter(s)). We are interested in
  \[ \hat{q}_0 = q_0(\hat{\theta}) = -2\log \frac{L(\mu = 0)}{L(\hat{\mu}, \hat{\theta})} = \max_{\theta} [q_0(\theta)] \]
  where \( \hat{\theta} \) is the \textit{global} maximum point.

- For which we want to know what is the p-value
  \[ \text{p-value} = P(\max_{\theta} [q_0(\theta)] \geq u), \quad u = Z^2 \]
The profile-likelihood test statistic
(with a nuisance parameter that is not defined under the Null hypothesis)

- Usually we only look for ‘positive’ signals
  (downward fluctuations of the BG are not considered as evidence against the BG)

\[
q_0(\theta) = \begin{cases} 
-2\log \frac{L(\mu = 0)}{L(\hat{\mu}, \theta)} & \hat{\mu} > 0 \\
0 & \hat{\mu} \leq 0
\end{cases}
\]

\[q_0(\theta) \] is ‘half chi^2’


The p-value just get divided by 1/2

- Or equivalently consider \( \hat{\mu} \) as a gaussian field

\[q_0(\theta) = \left( \frac{\hat{\mu}(\theta)}{\sigma} \right)^2 \]

( since \[q_0(\theta) = \left( \frac{\hat{\mu}(\theta)}{\sigma} \right)^2 \] )
In 1 dimension: points where the field values become larger than $u$ are called *upcrossings*.

The probability that the global maximum is above the level $u$ is called *exceedance probability*.

$$P(\max_\theta[q_0(\theta)] \geq u)$$
Random fields

- Fortunately, quite a lot of statistical literature on the properties of random fields


Applications in Cosmology, Brain mapping, Oceanography ...
The 1-dimensional case

For a chi\(^2\) random field, the expected number of upcrossings of a level \(u\) is given by: [Davies, 1987]

\[
E[N_u] = N_1 e^{-u/2}
\]

Note the inequality:

\[
E[N_u] \geq P(N_u > 0)
\]

To have the global maximum above a level \(u\):

- Either have at least one upcrossing \((N_u > 0)\) or have \(q_0 > u\) at the origin \((q_0(0) > u)\):

\[
P(\hat{q}_0 > u) \leq P(N_u > 0) + P(q_0(0) > u) \leq E[N_u] + P(q_0(0) > u)
\]

Becomes an equality for large \(u\)

[R.B. Davies, Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43 (1987)]
The 1-dimensional case

\[ E[N_u] = N_1 e^{-u/2} \]

The only unknown is \( N_1 \) which can be estimated from the average number of upcrossings at some low reference level.

\[ E[N_u] = N_1 e^{-u/2} \]
\[ E[N_{u_0}] = N_1 e^{-u_0/2} \]

\( N_1 = E[N_{u_0}] e^{u_0/2} \)

\[ E[N_u] = E[N_{u_0}] e^{(u_0-u)/2} \]

\[ p_{\text{global}} = E[N_{u_0}] e^{(u_0-u)/2} + p_{\text{local}} \]
Spin 0

Largest significance

$m_x \sim 750 \text{GeV}, \Gamma_x \sim 45 \text{GeV}(6)$

Local $Z = 3.9\sigma$

Any peak with $Z > 3.8\sigma$
with $m = 500$–$2000$ will draw our attention

$$P_{\text{global}}(u) \approx P_{\text{local}}(u) + E(n_{u_0})e^{\frac{u_0-u}{2}}$$

- $P_{\text{local}} = 5 \cdot 10^{-5}$
- $u_0 = 0$
- $n_{u_0} = 7 \pm 2.6$
- $u = Z^2 = 3.9^2 = 15.2$
- $P_{\text{global}} = 5 \cdot 10^{-5} + (7 \pm 2.6)e^{-15.2/2} = (2.2 - 4.8)10^{-3}$
- $Z_{\text{global}} \sim 2.7 \pm 0.1\sigma$

The LEE is even stronger when you consider another dimension (the width range (0-10%$m$) should also be taken into account.)
A real life example

\[ P(q_0 > u) \leq E[N_u] + P(q_0(0) > u) \]

\[ E[N_u] = N_1 e^{-u/2} \]

\[ N_1 \approx \langle N_{u_0} \rangle e^{u_0/2} \]

\[ P(q_0 > u) = N_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u) \]

\[ p_{\text{global}} = N_1 e^{-u/2} + p_{\text{local}} \]

\[ p_{\text{global}} = \langle N_{u_0} \rangle e^{-u_0/2} + p_{\text{local}} \]

\[ N_{u_0=0} = 9 \pm 3 \]

\[ p_{\text{global}} = 9 \cdot e^{-25/2} + O(10^{-7}) = 3.3 \cdot 10^{-5} \]

5\sigma \rightarrow 4\sigma \text{ trial#~100}
The Trial factor is given by

\[ u = Z^2 \]

\[ \text{trial } # = \frac{p_{global}}{p_{local}} \leq \frac{N_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)}{\frac{1}{2} P(\chi_1^2 > u)} \]

\[ N_1 \]

Is the number of independent search regions

\[ \text{trial } #(u >> 1) = 1 + \sqrt{\frac{\pi}{2} N_1 Z_{\text{fix}}} \approx \sqrt{\frac{\pi}{2} N_1 Z_{\text{fix}}} \]
Random fields (>1 D)

- The set of points where the field has values larger than some number $u$ is called the excursion set $A_u$ above the level $u$. 

Excursion set
Euler characteristic

- Number of disconnected components minus number of `holes`

\[ \phi = 1 \]

\[ \phi = 0 \]

\[ \phi = 2 \]

Excursion set
2-d example: search for neutrino sources (IceCube)

For a chi$^2$ field in 2 dimensions:

$$E[\varphi(A_u)] = \frac{1}{2} P(\chi^2 > u) + (N_1 + N_2 \sqrt{u}) e^{-u/2}$$

Estimate $E[\varphi]$ at two levels, e.g. 0 and 1, and solve for $N_1$ and $N_2$.

From 20 bkg. Simulations:

$$\langle \varphi_0 \rangle = 33.5 \pm 2$$
$$\langle \varphi_1 \rangle = 94.6 \pm 1.3$$

$$N_1 = 33 \pm 2$$
$$N_2 = 123 \pm 3$$
2-D example #2: resonance search with unknown width

- Gaussian signal on exponential background
- Toy model: 0 < m < 100, 2 < \sigma < 6
- Unbinned likelihood:

\[
\hat{q}_0
\]

\[
\sigma
\]

\[
m
\]

\[
\begin{align*}
L &= \prod_i \frac{N_s f_s(x_i) + N_b f_b(x_i)}{N_s + N_b} \times \text{Poiss}(N | N_s + N_b) \\
N_f &= N_s + N_b \\
f_s(x; m, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \\
f_b(x) &= ce^{-cx}
\end{align*}
\]
2-D example #2: resonance search with unknown width

\[ \mathcal{N}_1 = 4 \pm 0.2 \]
\[ \mathcal{N}_2 = 0.7 \pm 0.3 \]
The 750 GeV saga

2015

2D Scan

Largest significance

$m_x \sim 750\text{GeV}, \Gamma_x \sim 45\text{GeV}(6\%)$

Local $Z = 3.9\sigma$

$m=200-2000 \text{ GeV}$

\[ \frac{\Gamma_x}{m_x} = 0-10\% \]

Use toys or asymptotic formula from

arXiv:1105.4355

\[ Z_{local} = 3.9\sigma \]

\[ Z_{global} = 2.1\sigma \]

2.1σ is not something to write home about
CCGV Useful Formulae – The Bands

\[ \mu_{up}^{med} = \hat{\mu} + \sigma_{\mu_{up}}^{med} \phi^{-1}(1 - \alpha) \]

\[ \alpha = 0.05 \rightarrow \phi^{-1}(1 - \alpha) = \phi^{-1}(0.95) = 1.64 \]

\[ \mu_{up+N\sigma} = \hat{\mu} + \sigma_{up+N\sigma} \left( \phi^{-1}(1 - \alpha) + N \right) \]

\[ \sigma_{up+N\sigma}^{2} = \frac{\mu_{up+N\sigma}^{2}}{q_{up+N\sigma,A}} \]

Distribution of the upper limit with background only experiments

The Asimov data set is \( n=b \)

\[ \rightarrow \text{median upper limit} \]
Understanding the Brazil Plot

The expected 95% CL exclusion region covers the $m_H$ range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

- $\mu_{\text{up}} = \sigma(m_H)/\sigma_{SM}(m_H) < 1 \Rightarrow \sigma(m_H) < \sigma_{SM}(m_H) \Rightarrow \text{SM } m_H \text{ excluded}$

- The line $\mu_{\text{up}} = 1$ corresponds to $\text{CLs} = 5\%$ ($p'_{s} = 5\%$)

- The smaller $\mu_{\text{up}} < 1$ is, the exclusion of a SM Higgs is deeper $\Rightarrow p'_{s} < 5\%$, $p'_{s} = \text{CLs} \Rightarrow \text{CL} = 1 - p'_{s} > 95\%$