PREFIT-School
PRecision Effective Field Theory School
2-13 March 2020 at DESY, Hamburg

Fitting: Session 3

Andrea C. Marini, Nick Wardle

Assistants:

Lydia Brenner

Sebastian Wuchterl

Adinda de Wit

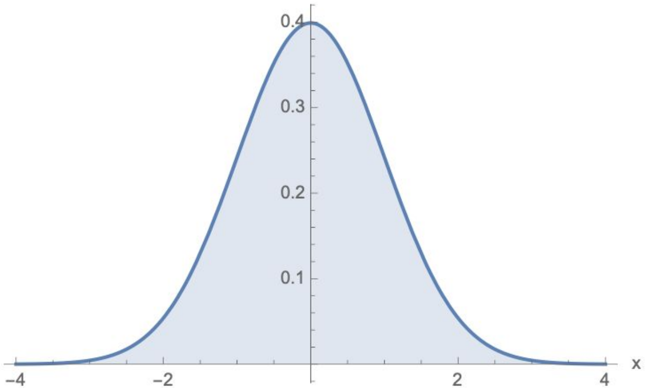
Statistical introduction

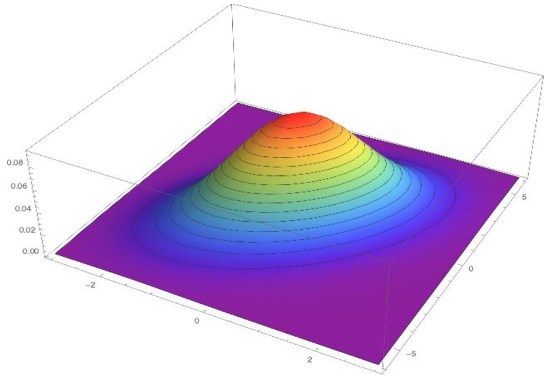
Normal Distribution

Continuos distributions

$$\mathcal{P}(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mathbb{E}[x] = \mu$$
$$\mathbb{V}[x] = \sigma^2$$





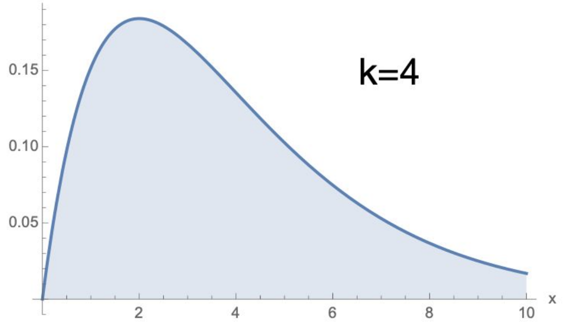
ChiSquare Distribution

- Chi Square distribution
  - sum of square normal distribution

$$X_i \sim \mathcal{N}(0,1)$$

$$Q = \sum_{i=0}^{k} X_i^2$$

k is the number
of degrees of freedom



$$\mathbb{E}[x] = k$$
$$\mathbb{V}[x] = 2k$$

Wilks' theorem

	68.3%	95%
1 ndf	1	3.84
2 ndf	2.23	5.99

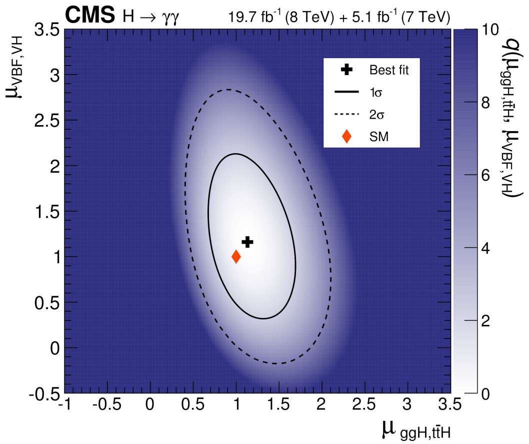
Likelihood ratio are often used in statistical tests:

a) H_0 is true

b) H_0 and H_1 are nested

c) Params for H_1, H_0 are well defined, and not on boundary

d) Data is asymptotic, i.e. the sample size approaches to infinity



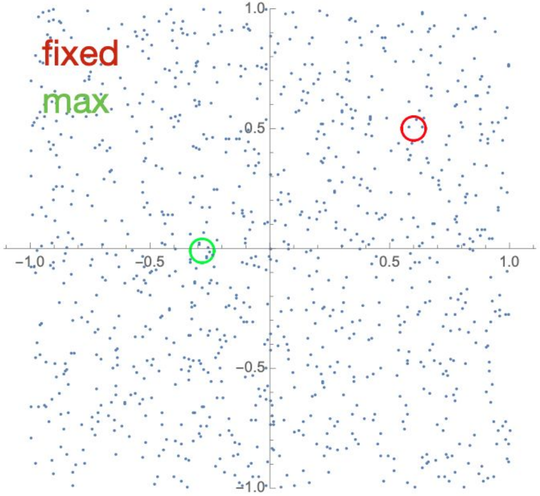
Then the $-2\log\Lambda$ is distributed as a χ^2 distribution with N degrees of freedom (where N is the difference in number of parameters between H_1 and H_0)

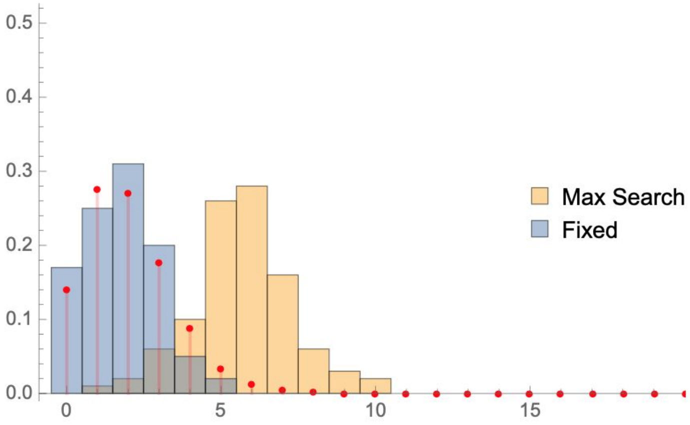
and $$\Lambda = \frac{L(\vec{\alpha})}{L(\hat{\vec{\alpha}})}$$

A word on LEE

- Look Elsewhere Effect (LEE) appears when you "cherry-pick" statistical results
  - Maximum finding, ecc..
- Very difficult to quantify
  - especially with nuisances, ecc..
  - use toys or LogL crossings

- 1000 points in $[-1,1]^2$
- $r_0 = 0.05$ -> Poisson with mean ~2
- how common is to observe 7?





Hands On

Objective of Hands On

What you will see today:

Material: https://github.com/amarini/Prefit2020

- MultiSignal Models
- Extract EFT couplings from STXS-like measurements

Part 1

Parametrization production

Look for available parametrization. We will use these ones (https://cds.cern.ch/record/2706103/files/HIG-19-005-pas.pdf). Stage 0.

STXS region (stage 0)	A_j
$gg \to H$	$8.73 \times 10^3\, c_G$
$qq \to Hqq$	$9.02\, c_{WW} + 0.6\, c_B - 0.797\, c_{HW} + 0.399\, c_A$
$qq \to H\ell\nu$	$42.5\, c_{WW} + 19.9\, c_{HW}$
$qq \to H\ell\ell$	$36.6\, c_{WW} + 10.5\, c_B + 15\, c_{HW} + 5.14\, c_A$
$gg/qq \to ttH$	$2.95\, c_u + 115\, c_G$

Table 10: A_j coefficients for the STXS stage 0 bins.

STXS region (stage 0)	B_{jk}
$gg \to H$	$1.95 \times 10^7\, {c_G}^2$
$qq \to Hqq$	$171\, {c_{WW}}^2 + 3.42\, {c_B}^2 + 114\, {c_{HW}}^2 + 0.874\, {c_A}^2 + 23.1\, c_{WW} c_B + 233\, c_{WW} c_{HW} + 6.22\, c_{WW} c_A + 15.3\, c_B c_{HW} + 2.02\, c_B c_A + 0.681\, c_{HW} c_A$
$qq \to H\ell\nu$	$912\, {c_{WW}}^2 + 558\, {c_{HW}}^2 + 1.3 \times 10^3\, c_{WW} c_{HW}$
$qq \to H\ell\ell$	$602\, {c_{WW}}^2 + 51.7\, {c_B}^2 + 321\, {c_{HW}}^2 + 10.7\, {c_A}^2 + 350\, c_{WW} c_B + 772\, c_{WW} c_{HW} + 102\, c_{WW} c_A + 227\, c_B c_{HW} + 31.4\, c_B c_A + 29.7\, c_{HW} c_A$
$gg/qq \to ttH$	$2.14\, {c_u}^2 + 6.13\, {c_{WW}}^2 + 1\, {c_B}^2 + 5.87\, {c_{HW}}^2 + 2.97 \times 10^4\, {c_G}^2 + 167\, c_u c_G - 0.31\, c_{WW} c_B + 11.9\, c_{WW} c_{HW} - 0.318\, c_B c_{HW}$

Table 11: B_{jk} coefficients for the STXS stage 0 bins.

Parametrization decay

And these ones (for the decay) https://cds.cern.ch/record/2673969

Partial width	$\sum_i A_i c_i$
$H \to b\bar{b}$	$-1.0\mathtt{cH} + 3.0\mathtt{cd}$
$H \to WW^* \to l\nu l\nu$	$10\mathtt{cWW} + 3.7\mathtt{cHW} + 2.2\mathtt{cpHL}$
$H \to ZZ^* \to 4l$	$55\mathtt{cWW} + 13\mathtt{cB} + 15\mathtt{cHW} + 4.6\mathtt{cHB} + 0.018c_\gamma + 2.0\mathtt{cHL} + 2.0\mathtt{cpHL} + 0.027\mathtt{cHe}$
$H \to \gamma\gamma$	$-5.8c'_\gamma$
$H \to \tau\tau$	$-1.0\mathtt{cH} + 3.0\mathtt{cl}$
$H \to gg$	$56c'_g$
$H \to$all	$0.0029\mathtt{cT} + 0.17\mathtt{cu} + 2.3\mathtt{cd} + 0.11\mathtt{cl} + 1.0\mathtt{cWW} + 0.023\mathtt{cB} + 0.37\mathtt{cHW}$ $+0.0079\mathtt{cHB} + 1.6c'_g + 0.0078\mathtt{cHQ} + 0.17\mathtt{cpHQ} + 0.0027\mathtt{cHu} + 0.057\mathtt{cpHL}$

Partial width	$\sum_{ij} B_{ij} c_i c_j$
$H \to b\bar{b}$	$0.25\mathtt{cH}^2 + 2.3\mathtt{cd}^2 + \mathtt{cH}(-1.5\mathtt{cd})$
$H \to c\bar{c}$	$0.25\mathtt{cH}^2 + 2.3\mathtt{cu}^2 + \mathtt{cH}(-1.5\mathtt{cu})$
$H \to \tau\tau$	$0.25\mathtt{cH}^2 + 2.3\mathtt{cl}^2 + \mathtt{cH}(-1.5\mathtt{cl})$
$H \to \gamma\gamma$	$8.4(c'^2_\gamma + c'^2_{\tilde\gamma})$
$H \to gg$	$790(c'^2_g + c'^2_{\tilde g})$
$H \to WW^* \to l\nu l\nu$	$0.25\mathtt{cH}^2 + 26\mathtt{cWW}^2 + 3.8\mathtt{cHW}^2 + 1.3\mathtt{cpHL}^2 + 0.32\mathtt{tcHW}^2$ $+\mathtt{cH}(-5.1\mathtt{cWW} - 1.9\mathtt{cHW} - 1.1\mathtt{cpHL}) + \mathtt{cWW}(19\mathtt{cHW}) + 12\mathtt{cpHL})$ $+\mathtt{cHW}(4.3\mathtt{cpHL})$
$H \to ZZ^* \to 4\ell$	$0.25\mathtt{cH}^2 + 4.0\mathtt{cT}^2 + 28\mathtt{cWW}^2 + 3.5\mathtt{cB}^2 + 2.2\mathtt{cHW}^2 + 0.20\mathtt{cHB}^2 + 1.8\mathtt{cHL}^2$ $+1.8\mathtt{cpHL}^2 + 0.43\mathtt{cHe}^2 + 0.14\mathtt{tcHW}^2 + \mathtt{cH}(2.0\mathtt{cT} - 5.1\mathtt{cWW} - 1.3\mathtt{cB}$ $-1.4\mathtt{cHW} - 0.43\mathtt{cHB} - 1.0\mathtt{cHL} - 1.0\mathtt{cpHL} + 0.43\mathtt{cHe}) + \mathtt{cT}(-21\mathtt{cWW}$ $-5.3\mathtt{cB} - 5.7\mathtt{cHW} - 1.7\mathtt{cHB} - 4.1\mathtt{cHL} - 4.1\mathtt{cpHL} + 1.7\mathtt{cHe}) + \mathtt{cWW}(10\mathtt{cB}$ $+15\mathtt{cHW} + 4.4\mathtt{cHB} + 12\mathtt{cHL} + 12\mathtt{cpHL} - 3.5\mathtt{cHe}) + \mathtt{cB}(3.8\mathtt{cHW} + 1.1\mathtt{cHB}$ $+0.052c'_\gamma + 1.1\mathtt{cHL} + 1.1\mathtt{cpHL} - 2.1\mathtt{cHe}) + \mathtt{cHW}(1.3\mathtt{cHB} + 3.0\mathtt{cHL}$ $+3.0\mathtt{cpHL} - 1.3\mathtt{cHe}) + \mathtt{cHB}(0.91\mathtt{cHL} + 0.91\mathtt{cpHL} - 0.39\mathtt{cHe})$ $+\mathtt{cHL}(3.5\mathtt{cpHL} - 0.13\mathtt{cHe}) + \mathtt{cpHL}(-0.13\mathtt{cHe}) + 0.081\mathtt{tcHW}(\mathtt{tcHB})$
$H \to$all	$0.24\mathtt{cH}^2 + 0.037\mathtt{cT}^2 + 0.13\mathtt{cu}^2 + 1.7\mathtt{cd}^2 + 0.084\mathtt{cl}^2 + 2.6\mathtt{cWW}^2 + 4.7\mathtt{cHW}^2$ $+4.3\mathtt{cHB}^2 + 23c'^2_g + 0.09\mathtt{cpHQ}^2 + 0.066\mathtt{cHud}^2 + 0.027\mathtt{cpHL}^2 + 4.3\mathtt{tcHW}^2$ $+4.3\mathtt{tcHB}^2 + 23c'^2_{\tilde g} + \mathtt{cH}(-0.086\mathtt{cu} - 1.2\mathtt{cd} - 0.056\mathtt{cl} - 0.51\mathtt{cWW}$ $-0.18\mathtt{cHW} - 0.083\mathtt{cpHQ} - 0.029\mathtt{cpHL}) + \mathtt{cT}(-0.19\mathtt{cWW} - 0.046\mathtt{cB}$ $-0.051\mathtt{cHW} - 0.027\mathtt{cpHQ}) + \mathtt{cWW}(0.11\mathtt{cB} + 1.9\mathtt{cHW} + 0.04\mathtt{cHB} + 0.86\mathtt{cpHQ}$ $+0.29\mathtt{cpHL}) + \mathtt{cHW}(0.03\mathtt{cB} - 8.6\mathtt{cHB} + 0.1c_\gamma + 0.31\mathtt{cpHQ} + 0.11\mathtt{cpHL})$ $+\mathtt{cHB}(-0.1c_\gamma) + \mathtt{tcHW}(-8.6\mathtt{tcHB} + 0.1c'_{\tilde\gamma}) + \mathtt{tcHB}(-0.10c'_{\tilde\gamma})$

Prepare measurements

You have 3 stage 0 categories and one decay:

ggH, qqH, ttH, x Hgg

```
OBJ: TH1F     data_ggH_hgg__mgg      Histogram of data_ggH_hgg__mgg : 0 at: 0x6adb920
OBJ: TH1F     data_qqH_hgg__mgg      Histogram of data_qqH_hgg__mgg : 0 at: 0x6c56870
OBJ: TH1F     data_ttH_hgg__mgg      Histogram of data_ttH_hgg__mgg : 0 at: 0x6da0750
OBJ: TH2D     Response       sig : 0 at: 0x6c602f0
```

The data are in the input file (`Session 3/inputs_session3.root`).

They are `TH1F`s (you need to convert them into `RooDataHists`) eg.

In ROOT

```
TFile *fi = TFile::Open("inputs_session3.root");
RooRealVar x("x","x",110,160);
RooDataHist data1("data1","ggH_tagged_data",RooArgList(x),(TH1*)fi.Get("data_ggH_hgg__mgg"));
```

Or in PyROOT

```
import ROOT
fi = ROOT.TFile.Open("inputs_session3.root")
x = ROOT.RooRealVar("x","x",110,160)
data1 = ROOT.RooDataHist("data1","ggH_tagged_data",ROOT.RooArgList(x),fi.Get("data_ggH_hgg__mgg"))
```

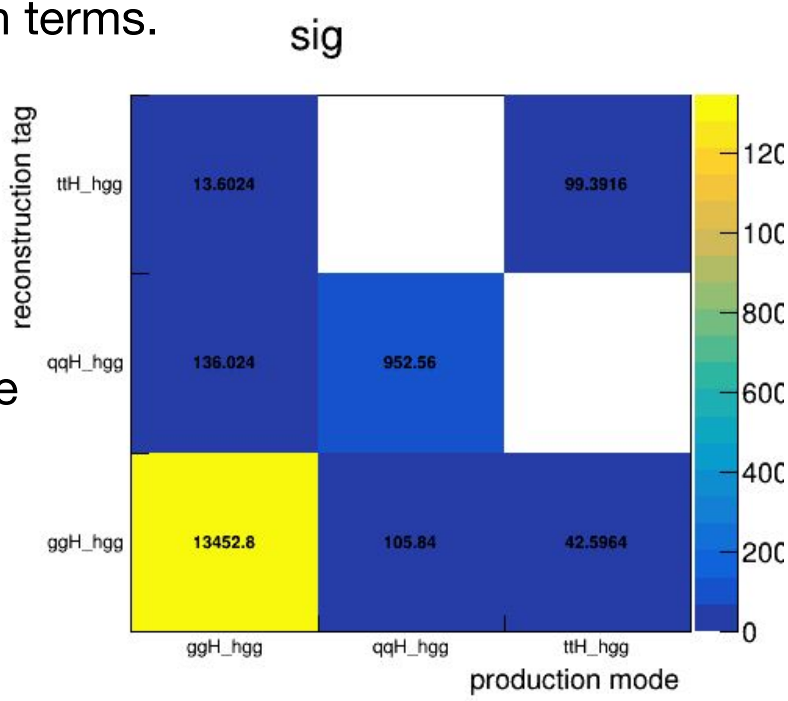
Prepare measurements

The expected signal events are written in the Response 2D Histogram with the relative confusion terms.

The signal is a gaussian with width s = 1.5GeV

The background is exponential

You can use what you learned in Session 1 to make the pdfs for the signal and background in each category.



Assignment: Part 1

Measure the Stage 0 cross section for ggH, qqH and ttH.

Use that the injected luminosity is L=140 fb^{-1}

Derive the error and the covariance matrix of the fit.

Part 2

Measure directly the EFT params

If we float more than one EFT Parameter at the time, we need to check if we have a simultaneous sensitivity to all of them.

The production scaling depends on c_G, c_{WW}, c_B, c_{HW}, c_A, ...

STXS region (stage 0)	A_j
gg → H	$8.73 \times 10^3\, c_G$
qq → Hqq	$9.02\, c_{WW} + 0.6\, c_B - 0.797\, c_{HW} + 0.399\, c_A$
gg/qq → ttH	$2.95\, c_u + 115\, c_G$

Table 10: A_j coefficients for the STXS stage 0 bins.

STXS region (stage 0)	B_{jk}
gg → H	$1.95 \times 10^7\, {c_G}^2$
qq → Hqq	$171\, {c_{WW}}^2 + 3.42\, {c_B}^2 + 114\, {c_{HW}}^2 + 0.874\, {c_A}^2 + 23.1\, c_{WW} c_B + 233\, c_{WW} c_{HW} + 6.22\, c_{WW} c_A + 15.3\, c_B c_{HW} + 2.02\, c_B c_A + 0.681\, c_{HW} c_A$
gg/qq → ttH	$2.14\, {c_u}^2 + 6.13\, {c_{WW}}^2 + 1\, {c_B}^2 + 5.87\, {c_{HW}}^2 + 2.97 \times 10^4\, {c_G}^2 + 167\, c_u c_G - 0.31\, c_{WW} c_B + 11.9\, c_{WW} c_{HW} - 0.318\, c_B c_{HW}$

Table 11: B_{jk} coefficients for the STXS stage 0 bins.

Measure directly the EFT params

Start from these scaling equations:

ggH : $1 + 8.73c'_G + 19.5(c'_G)^2$

qqH : $1 + 0.6c_B - 0.797c_{HW} + 3.42c_B^2 + 114c_{HW}^2 + 15.3c_{HW}c_B$

ttH : $1 + 0.115c'_G + c_B^2 + 5.87c_{HW}^2 + 0.0297(c'_G)^2 - 0.318c_Bc_{HW}$

And derive the equivalent functions in terms of c'_G and c_{HW}-c_B (assuming $c_{HW}+c_B=0$)

Now remake your signal models such that the scaling functions for the ggH, qqH and ttH processes are the ones you just derived

*Note that we have included the large factors of 10 inside the coefficient c_G and called it c'_G simply to avoid fitting with very small numbers

Part 3

Prop. the Cov matrix to the EFT

Use the covariance matrix and the measurements you obtained in part 1 to derive the constraints on the parameter c_G.

Limitations:

- Gaussian approximation
- Nuisance parameters and correlation among them are (partially) neglected accordingly to what available.

Conclusions

Summary & Conclusions

We covered …

- Some basic RooFit
    - Object creation/manipulation, pdfs and toy-generation, likelihood construction and minimization
- Simultaneous likelihoods
    - Multiple bins / multiple categories of data → additional constraints on physics parameters from including additional data in the LH
- Physics parameter determination
    - Multiple signal processes can contribute to multiple categories (regions) in data → unfolding allows to extract those contributions (cross-sections)
    - Being able to determine differences from expected contributions can be used to constraint EFT coefficients from the experimental data

Thanks!