Statistical Analysis Walk-Through: Toy Search for the 125 GeV Higgs in the 4-Muon Final State

Håkon Fossheim

Universitetet i Oslo

09.09.2020

Overview

1. Theoretical Background
2. Finding the Optimal Mass Window
3. Estimating Scalefactors via Maximum Likelihood
4. The Optimal test-statistic
5. Discovery and Exclusion Aimed

Theoretical Background

Total number of particle interactions in our experiment is given by

$$N = \sigma \int dt \mathcal{L} \tag{1}$$

Where $\mathcal{L}$ is the instantaneous luminosity. Assuming this is constant during our data acquisition period, we have

$$N = \sigma \mathcal{L} \Delta t = \sigma L \tag{2}$$

Now, each particle interaction has a fixed probability, p_i, to end up in bin i in our histogram. The expected number of events in this bin, is thus

$$\lambda_i = N p_i \tag{3}$$

We can show that the probability distributions in any given bin is, to a very good approximation, given by the Poisson distribution, $P(n_i; \lambda_i)$. Here λ_i is the expectation value of whichever hypothesis we're considering. If we expect b_i and $s_i + b_i$ events in this bin under H_0 and H_1 respectively, We get

Probability of observing n events under H_0:

$$\mathrm{P}_{\mathrm{H}_0}(n; b) = \frac{b^n}{n!} e^{-b} \tag{4}$$

Probability of observing n events under H_1:

$$\mathrm{P}_{\mathrm{H}_1}(n; s+b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \tag{5}$$

Where we dropped the index for readability.

When specifying an alternative hypothesis H_1, we are also postulating a specific cross section for the process in question. We can therefore expect the ratio $\frac{s}{b}$ for this bin to be constant as a function of the total number of interactions, N. Leaving the index implicit, we get

$$\frac{s}{b} = k, \qquad \text{with k} = \text{const} \tag{6}$$

and so

$$\langle \mathrm{H}_0 \rangle = b \tag{7}$$
$$\langle \mathrm{H}_1 \rangle = b(1 + k) \tag{8}$$

From Eq.(3) and Eq.(2), we have

$$b = Np_b = \sigma L \tag{9}$$

- We therefore expect H_0 and H_1 to **remain separated by a fixed distance** (in units of b) as L increases

Furthermore,

- A Poisson distribution **becomes more localized** w.r.t. λ as λ increases

This can be seen by considering the distance between the mean and and one standard deviation away from it:

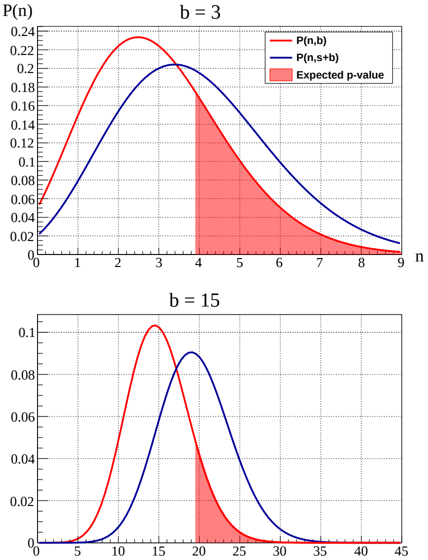
$$\frac{(\mu+\sigma)-\mu}{\mu} = \frac{\sigma}{\mu} = \frac{\sqrt{\lambda}}{\lambda} \tag{10}$$

Which approaches 0 as $\lambda \to \infty$.

- Thus, as N increases, $\langle \mathrm{H}_0 \rangle$ and $\langle \mathrm{H}_1 \rangle$ should remain separated by a "fixed" distance, while they become more "localized".

The expected P-value should therefore decrease as we increase N through increasing the luminosity, which again results in a higher expected significance, Z_{exp}.

- As we will soon see, Z_{exp} should therefore increase monotonically with L for a given bin/mass-window.


b = 3
P(n)
n
P(n,b)
P(n,s+b)
Expected p-value
b = 7
b = 15
b = 50

Finding the Optimal Mass Window

- In an ideal world, we would be able to filter out all non-Higgs events
- However, some events are too similar and can't be filtered out
- We therefore make one last cut, namely the mass window we wish to consider

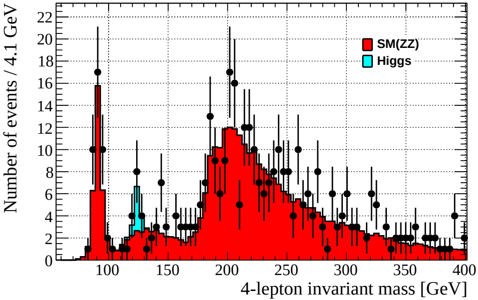


Figure (3)

- The optimal mass window is that which gives the highest expected significance, as this maximizes the sensitivity of our hypothesis test
- One might initially think that the optimal cut would correspond to the **bin at the signal peak**, as this is where $\frac{\mathbf{s}}{\mathbf{b}}$ **is largest** and thus the mean of the two distributions will maximally separated (in units of b).
- However, this would give a very low luminosity, and the distributions would be quite smeared out and overlapping.
- Increasing the luminosity by increasing the full-width around the mass window gives more "localized" distributions at the cost of decreasing $\frac{s}{b}$.

- If the fullwidth is very small
  - $\frac{s}{b}$ large, but L too small
- If fullwidth is very large
  - L large, but $\frac{s}{b}$ too small

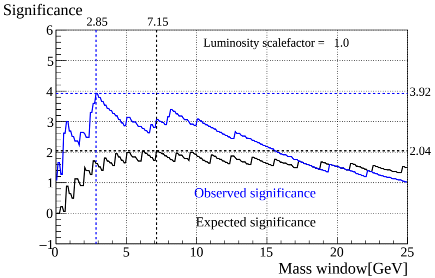
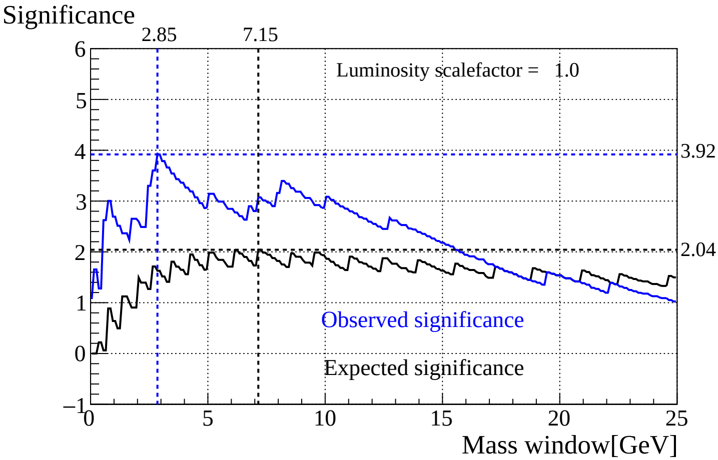


Figure (4)

There is therefore a "tug war" between the ratio $\frac{s}{b}$ and L. We can find the compromise of L and $\frac{s}{b}$ which gives our hypothesis test the best sensitivity by plotting the expected significance vs. the window full-width.



	Expected	Observed
Optimal width	7.15 GeV	2.85 GeV
Corresponding Z	2.04 σ	3.92 σ

Table (1) Optimal mass windows and their corresponding significances.

We can increase the luminosity of our **dataset** to see what significance one could expect to achieve with more data

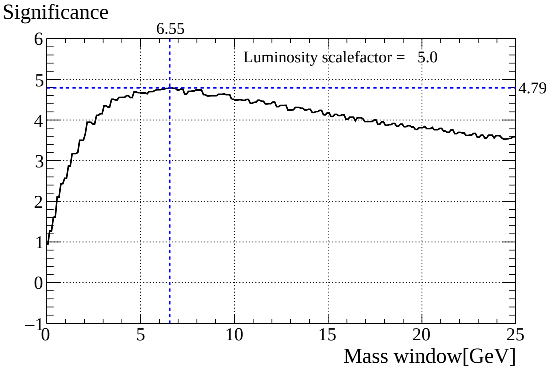


Figure (5)

- 5 times higher luminosity yields a test with expected significance of $Z_{exp} = 4.79$. Almost at the 5σ requirement for discovery!

We therefore realize that if we are to **expect to be able to make a discovery**, we need to **increase the size of our datasample**, i.e. our luminosity. We can find how much we need to increase it by plotting the expected significance against the luminosity scalefactor (our dataset currently has a scalefactor of 1 by definiton).

- Need to increase our luminosity by a factor of 5.38 to expect to be able to meet the 5 σ discovery requirement
- As predicted, the expected significance increases monotonically with luminosity

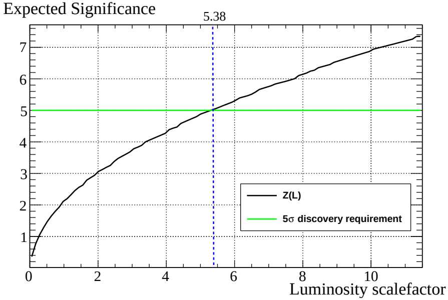


Figure (6)

Estimating MC Scalefactors via Maximum Likelihood

In order to determine the best fit of the simulated background and signal distributions, we turn to the likelihood function.

- Measures how probable the observed data, **n**, are for various value of the shape parameters α and μ

Is given by the product of the PMF of each bin, namely:

$$L(\alpha, \mu; \mathbf{n}) = \prod_{i=1}^{B} P(n_i; \alpha b + \mu s) \tag{11}$$

- Measures how probable the observed data, **n**, are for various values of the shape parameters α and μ.
- α and μ globally scale the background and signal+background estimates, respectively

Single parameter case, scaling the background only

$$\frac{\partial \ln L(\alpha, \mu = 0; \mathbf{n})}{\partial \alpha} = 0 \tag{12}$$

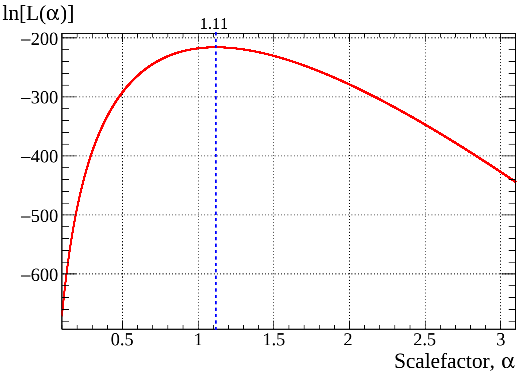


Figure (7)

The likelihood function will be Gaussian distributed in the limit of an infinite datasample,

$$L(\alpha) = Ce^{-\frac{(\alpha-\alpha_{max})^2}{2\sigma^2_{\alpha_{max}}}}$$

$$\ln L(\alpha) = -\frac{(\alpha-\alpha_{max})^2}{2\sigma^2_{\alpha_{max}}} + \ln C$$

Using $C = L(\alpha_{max})$ gives

$$\ln L(\alpha_{max}) - \ln L(\alpha) = \frac{1}{2}\frac{(\alpha-\alpha_{max})^2}{\sigma^2_{\alpha_{max}}}$$

$$\ln L(\alpha_{max}) - \ln L(\alpha) = \frac{1}{2} \iff |\alpha-\alpha_{max}| = \sigma_{\alpha_{max}}$$

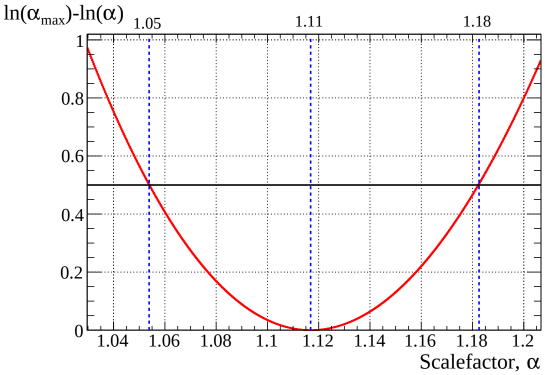


Figure (8)

$$\alpha_{max}{}^{\alpha_{up}-\alpha_{max}}_{\alpha_{max}-\alpha_{low}} = \alpha_{max}{}^{+\Delta\alpha}_{-\Delta\alpha} = 1.11^{+0.07}_{-0.06}$$

$$b_{scaled}{}^{+\Delta b}_{-\Delta b} = 1.11^{+0.07}_{-0.06} \cdot 4.83 = 5.39^{+0.32}_{-0.30}$$

Can use this to calculate $\langle Z_{rescaled} \rangle = 2.027$. Smaller than the previous $\langle Z_{unscaled} \rangle = 2.05$

Two parameter case, scaling background and signal+background

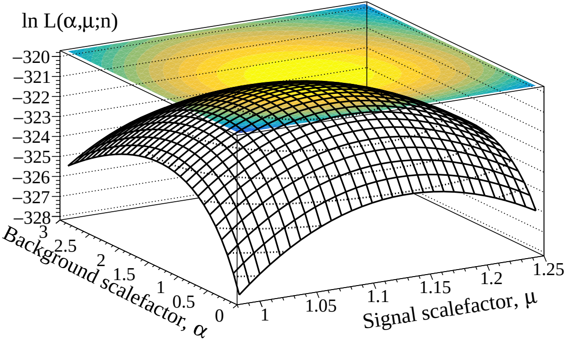


Figure (9)

$$\alpha_{max} = 1.11, \ \ \mu_{max} = 1.28 \tag{13}$$

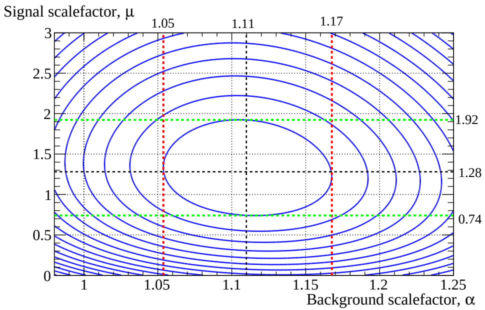


Figure (10)

- Had to restrict the range of α due to its low variance

$$\alpha_{max} = 1.11^{+0.058}_{-0.056}, \quad \mu_{max} = 1.28^{+0.64}_{-0.54} \tag{14}$$

The optimal test-statistic

- Reduces our dataset to a single number. A common choice:

$$t(\mathbf{n},\mu) = -2\ln\lambda = -2\ln\left\{\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}\right\}$$

- We will use a slightly different version:

$$t(\mathbf{n}) = -2\ln\lambda = -2\ln\left\{\frac{L(\mathbf{n};\mu=1,\hat{\hat{\alpha}}_{\mu=1})}{L(\mathbf{n};\mu=0,\hat{\hat{\alpha}}_{\mu=0})}\right\} \quad (15)$$

Yields the greatest separation of the two hypotheses one is testing for.

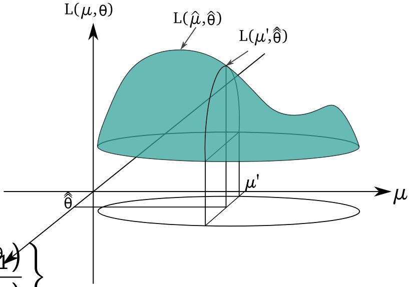
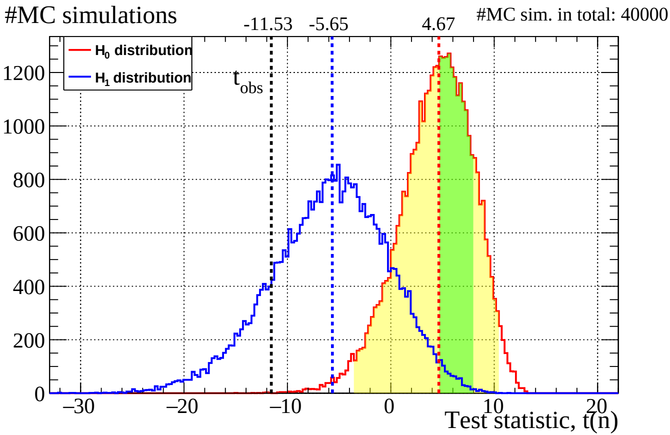


Figure (11)

- Finding the distributions under $t(\mathbf{n})$ analytically is non-trivial

Instead, we generate them by use of *pseudo-experiments*:

- For each experiment:
  - Set each bin to a random $P(n_i; b)$ or $P(n_i; s+b)$ distributed value
  - Calculate $t(\mathbf{n})$ and increment the corresponding bin in the following figure by one:



#MC simulations
#MC sim. in total: 40000
-11.53
-5.65
4.67
H0 distribution
H1 distribution
tobs
1200
1000
800
600
400
200
0
–30
–20
–10
0
10
20
Test statistic, t(n)

Figure (12)

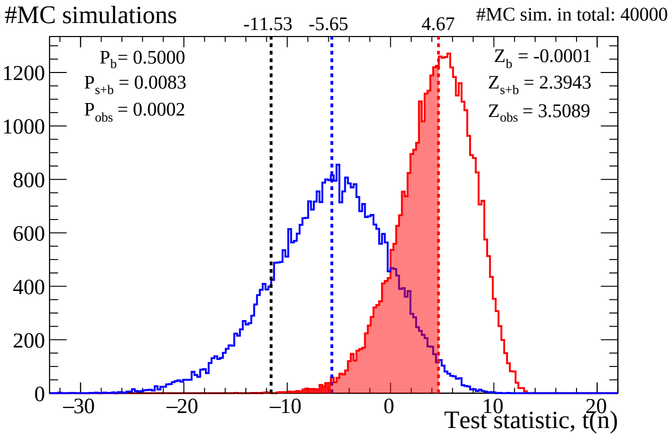

#MC simulations
#MC sim. in total: 40000
-11.53
-5.65
4.67
P_b = 0.5000
P_s+b = 0.0083
P_obs = 0.0002
Z_b = -0.0001
Z_s+b = 2.3943
Z_obs = 3.5089
1200
1000
800
600
400
200
0
–30
–20
–10
0
10
20
Test statistic, t(n)

Figure (13)

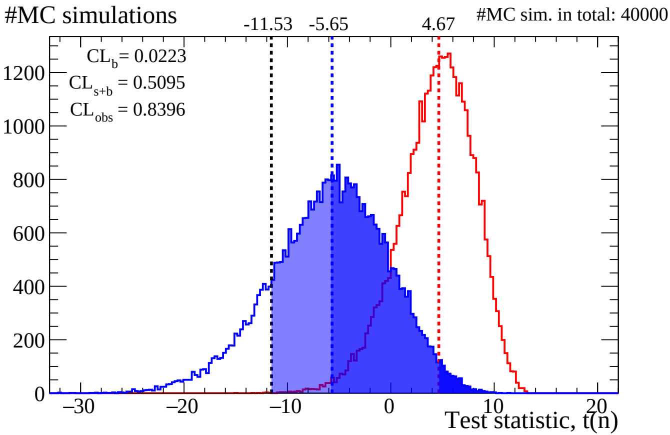

#MC simulations
#MC sim. in total: 40000
-11.53
-5.65
4.67
CL_b = 0.0223
CL_s+b = 0.5095
CL_obs = 0.8396
1200
1000
800
600
400
200
0
–30
–20
–10
0
10
20
Test statistic, t(n)

Figure (14)

$\frac{s}{b}$ is fixed for a given alternative hypothesis, since we are testing for a specific σ. Increasing σ corresponds to increasing this s with b fixed, and H_0 gets shifted left. This can be achieved by increasing the signal scalefactor μ.

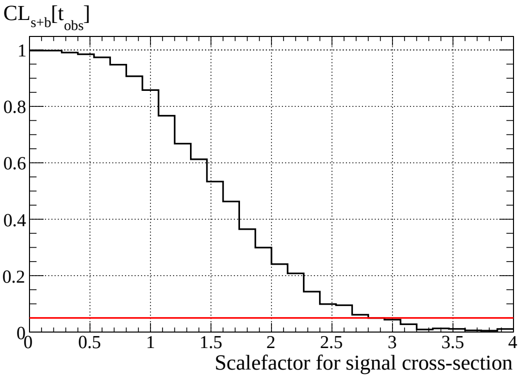
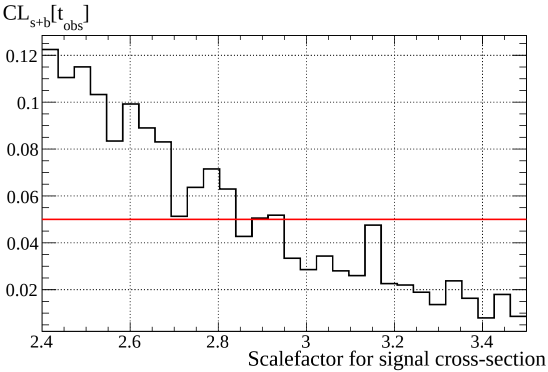


Figure (15)



CLs+b[tobs]
0.12
0.1
0.08
0.06
0.04
0.02
2.4
2.6
2.8
3
3.2
3.4
Scalefactor for signal cross-section

Figure (16)