

Search and Discovery Statistics in HEP Lecture 2.1

Eilam Gross, Weizmann Institute of Science

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

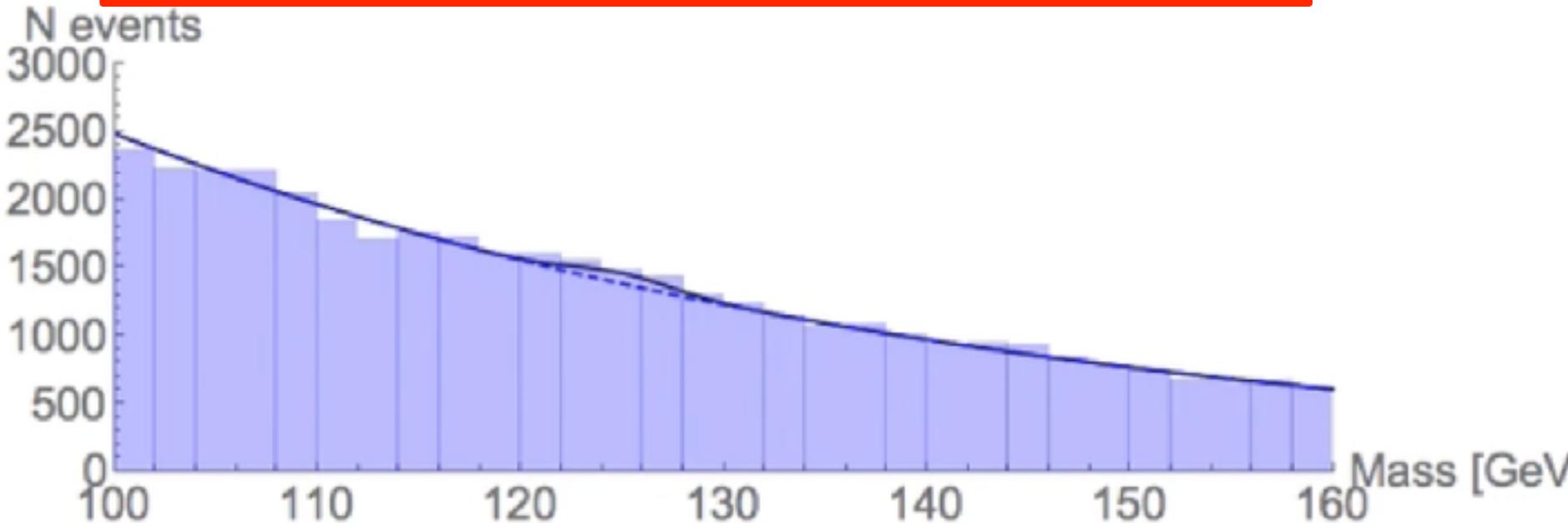
Louis Lyons, Alex Read, Bob Cousins Glen Cowan ,Kyle Cranmer
Ofer Vitells & Jonathan Shlomi



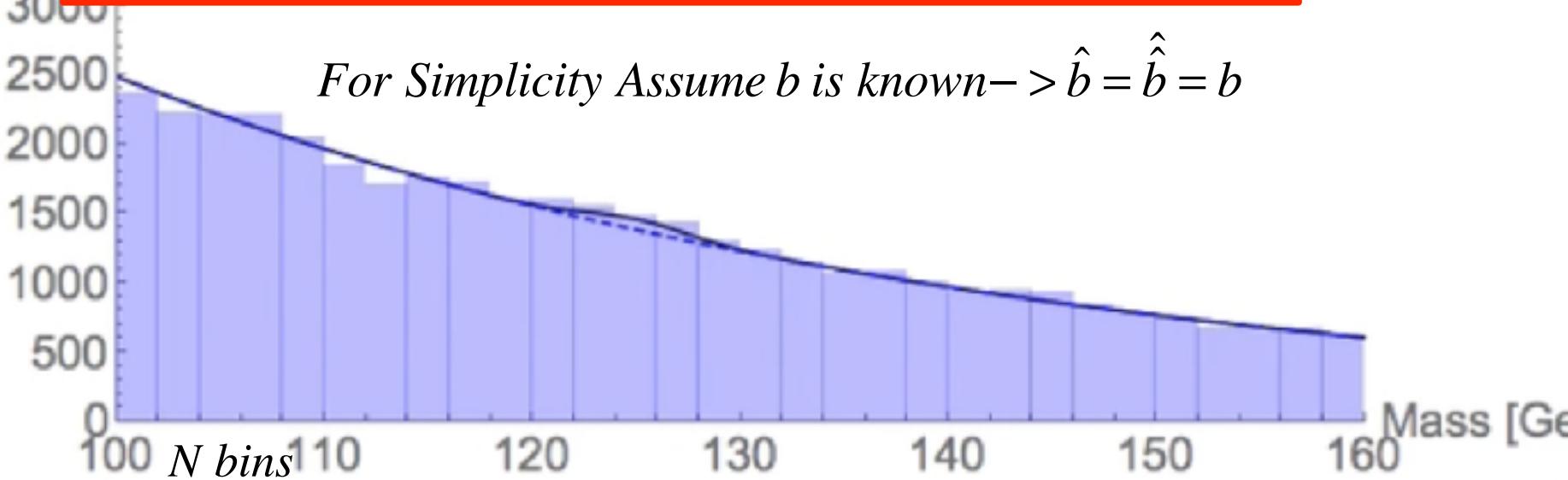
Bump Hunt ($H \rightarrow \text{GammaGamma}$)

- Find the expected sensitivity for Higgs discovery with a Luminosity equivalent to 70000 events
- There are 70000 entries in the histogram

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \quad \lambda(0) = \frac{L(\mu=0)}{L(\hat{\mu})} = \frac{L(\hat{b}_{\mu=0})}{L(\hat{\mu}s + \hat{b})} = \frac{L(\hat{b}_{\mu=0})}{L(\hat{s} + \hat{b})}$$



$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \quad \lambda(0) = \frac{L(\mu=0)}{L(\hat{\mu})} = \frac{L(\hat{b}_{\mu=0})}{L(\hat{\mu}s + \hat{b})} = \frac{L(\hat{b}_{\mu=0})}{L(\hat{s} + \hat{b})}$$

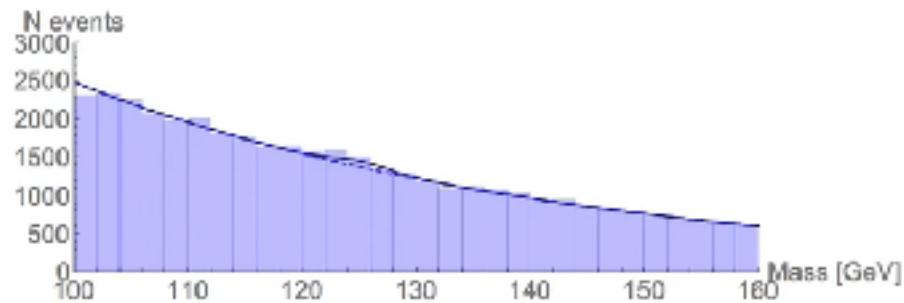
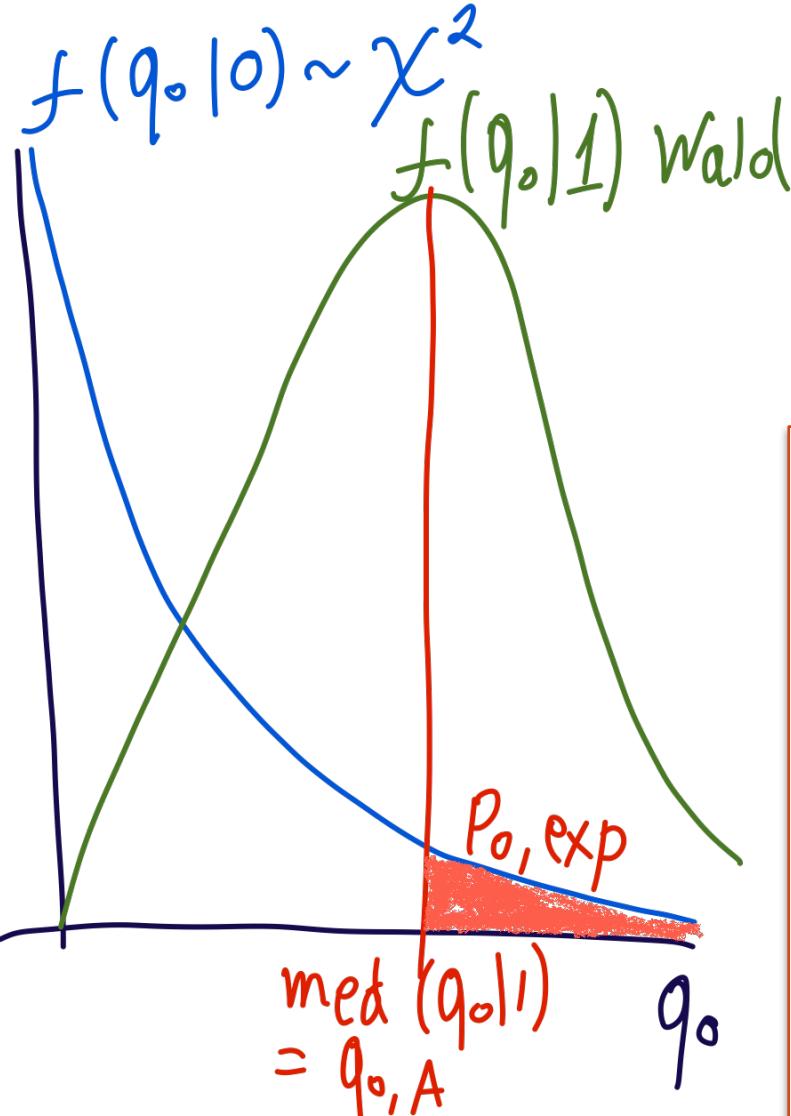


$$L(\mu) = \prod_{i=1,N} L(n_i | \mu s_i + b_i) = \prod_{i=1,N} \frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{n_i}}{n_i!}$$

$$L(0) = \prod_{i=1,N} L(n_i | b_i) = \prod_{i=1,N} \frac{e^{-(b_i)} (b_i)^{n_i}}{n_i!}$$



Sensitivity for Discovery



To find the expected sensitivity, calculate

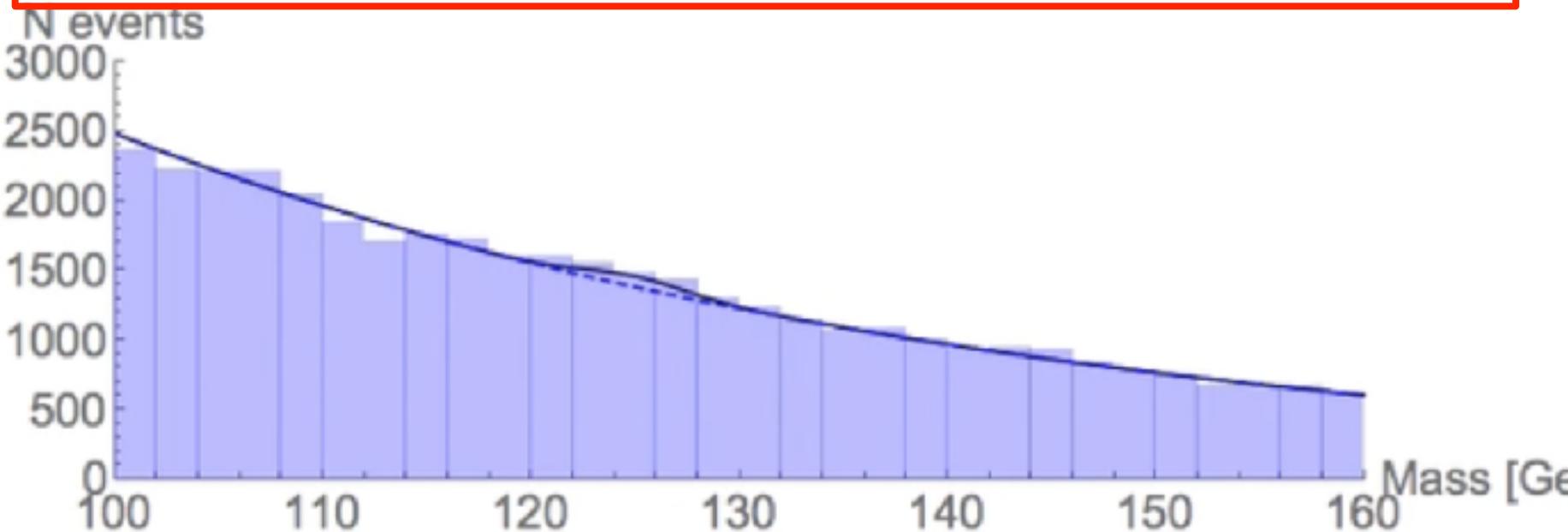
$$P_{0,\text{exp}} = \int_{q_0,A}^{\infty} f(q_0|0) dq_0$$

If $P_{0,\text{exp}} \gtrsim 2.9 \cdot 10^{-7}$ there is a sensitivity for 5 σ discovery



Asimov Data Set

$$q_{0,A} = \begin{cases} -2 \ln \lambda_A(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \quad \lambda_A(0) = \frac{L(\mu=0)}{L(\hat{\mu}=1)} = \frac{L(b)}{L(s+b)} = \frac{Pois(n=s+b|b)}{Pois(n=s+b|s+b)}$$



Asimov Data Set

$$q_{0,A} = \begin{cases} -2 \ln \lambda_A(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} \quad \lambda_A(0) = \frac{L(\mu=0)}{L(\hat{\mu}=1)} = \frac{L(b)}{L(s+b)} = \frac{Pois(n=s+b|b)}{Pois(n=s+b|s+b)}$$

N bins

$$L_A(0) = \prod_{i=1,N} L\left(n_i = s_i + b_i | b_i\right) = \prod_{i=1,N} \frac{e^{-(b_i)} (b_i)^{s_i+b_i}}{(s_i + b_i)!}$$

$$L_A(\hat{\mu}=1) = \prod_{i=1,N} L\left(n_i = s_i + b_i | s_i + b_i\right) = \prod_{i=1,N} \frac{e^{-(s_i+b_i)} (s_i + b_i)^{s_i+b_i}}{(s_i + b_i)!}$$

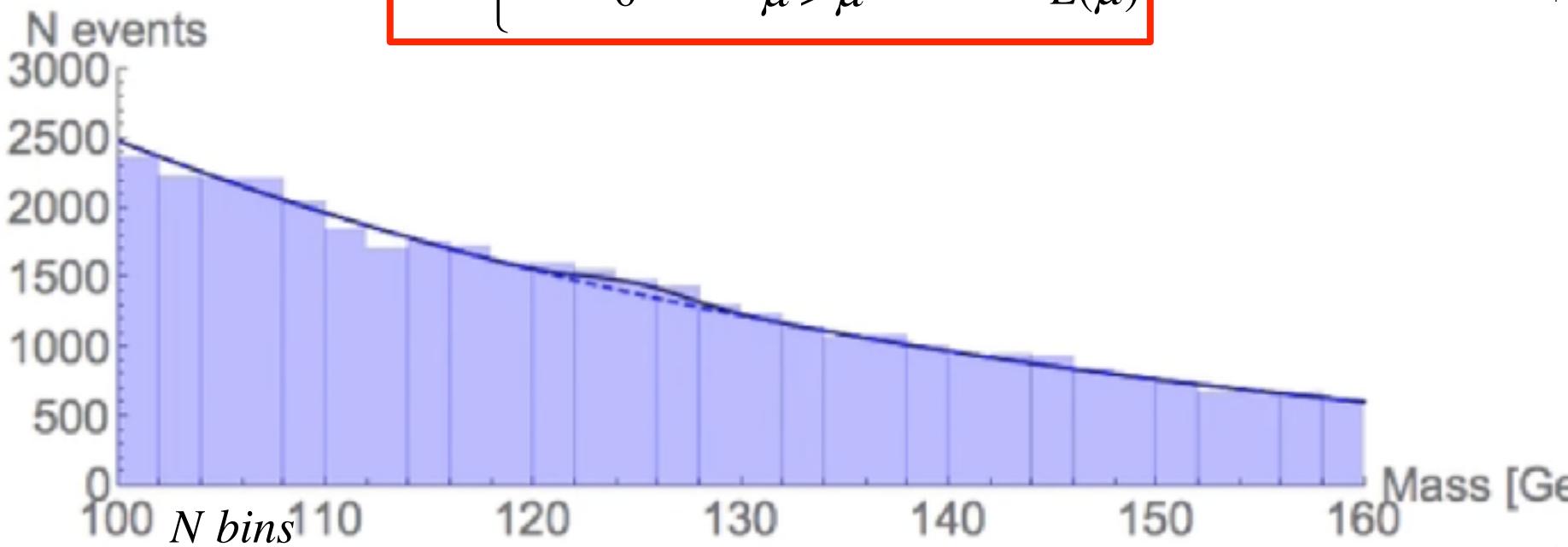
$$q_{0,A} = -2 \ln \frac{L_A(0)}{L_A(\hat{\mu}=1)} = -2 \ln \prod_{i=1,N} \frac{(s_i + b_i)!}{\frac{e^{-(s_i+b_i)} (s_i + b_i)^{s_i+b_i}}{(s_i + b_i)!}} = -2 \ln \prod_{i=1,N} \frac{e^{-(b_i)} (b_i)^{s_i+b_i}}{e^{-(s_i+b_i)} (s_i + b_i)^{s_i+b_i}}$$

$$q_{0,A} = -2 \ln \prod_{i=1,N} \frac{e^{-(b_i)} (b_i)^{s_i+b_i}}{e^{-(s_i+b_i)} (s_i + b_i)^{s_i+b_i}} = -2 \ln \prod_{i=1,N} \frac{e^{s_i}}{\left(1 + \frac{s_i}{b_i}\right)^{s_i+b_i}}$$



Exclusion

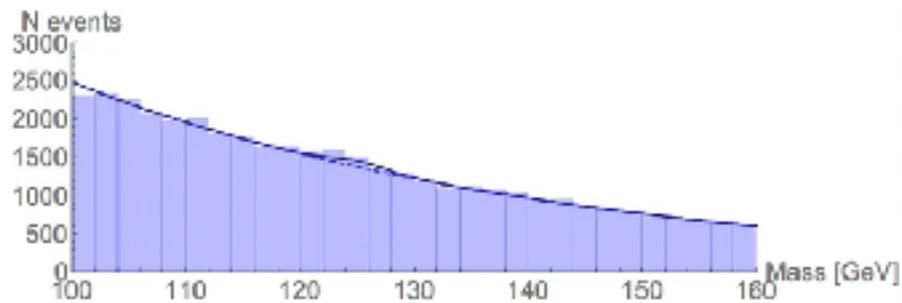
$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad \lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})}$$



$$L(\mu) = \prod_{i=1,N} L(n_i | \mu s_i + b_i) = \prod_{i=1,N} \frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{n_i}}{n_i!}$$



Sensitivity for Exclusion



To find the expected exclusion sensitivity calculate

$$P_{\mu, \text{exp}} = \int_{q_{\mu,A}}^{\infty} f(q_\mu | \mu) dq_\mu$$

If $P_{\mu, \text{exp}} \leq 5\%$ there is a sensitivity for exclusion



Asimov Data Set

$$q_{\mu,A} = \begin{cases} -2 \ln \lambda_A(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad \lambda_A(\mu) = \frac{L(\mu)}{L(\hat{\mu}=0)} = \frac{L(\mu s + b)}{L(b)} = \frac{Pois(n=b | \mu s + b)}{Pois(n=b | b)}$$

N bins

$$L_A(\mu) = \prod_{i=1,N} L(n_i = b_i | \mu s_i + b_i) = \prod_{i=1,N} \frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{b_i}}{(b_i)!}$$

$$L_A(\hat{\mu} = 0) = \prod_{i=1,N} L(n_i = b_i | b_i) = \prod_{i=1,N} \frac{e^{-(b_i)} (b_i)^{b_i}}{(b_i)!}$$

$$\frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{b_i}}{(b_i)!}$$

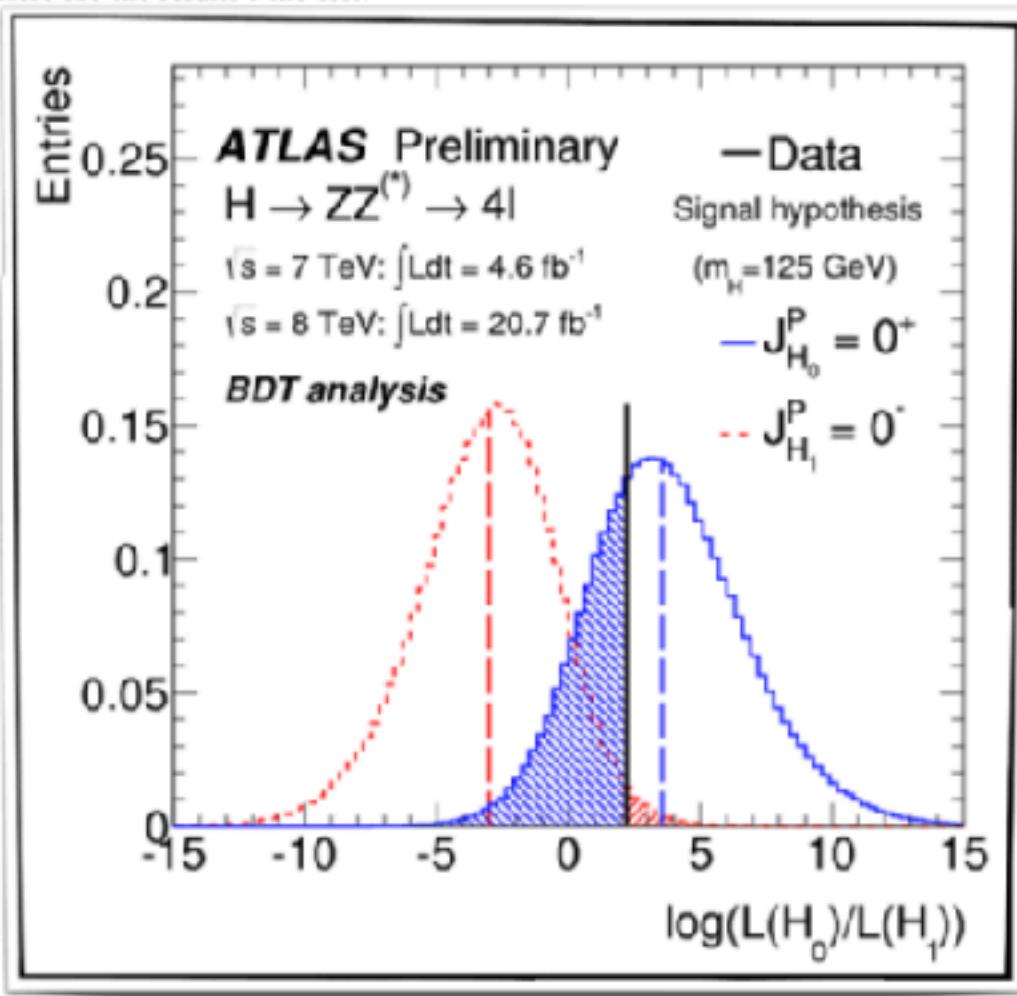
$$q_{\mu,A} = -2 \ln \frac{L_A(\mu)}{L_A(\hat{\mu} = 0)} = -2 \ln \prod_{i=1,N} \frac{\frac{(b_i)!}{e^{-(b_i)} (b_i)^{b_i}}}{\frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{b_i}}{(b_i)!}} = -2 \ln \prod_{i=1,N} \frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{b_i}}{e^{-(b_i)} (b_i)^{b_i}}$$

$$q_{\mu,A} = -2 \ln \prod_{i=1,N} \frac{e^{-(\mu s_i + b_i)} (\mu s_i + b_i)^{b_i}}{e^{-(b_i)} (b_i)^{b_i}} = -2 \ln \prod_{i=1,N} e^{-\mu s_i} \left(1 + \frac{\mu s_i}{b_i} \right)^{b_i}$$



Here we test the $J^P = 0^+$ hypothesis against the $J^P = 0^-$ hypothesis.

Here are the results of the test:



1. The test statistics is given by the Nyman Pearson ine, i.e. $q = -2 \ln \frac{L(H_x)}{L(H_y)}$, what is x and y? Explain.

2. Mark on the plot the various p-values. i.e. p_{H_1} (expected), p_{H_1} (observed), p_{H_0} (expected), p_{H_0} (observed)

3. Which p values correspond to 31%, 0.37% and 1.5%?

4. Calculate the modified CLs value of p_{H_1} .

5. At which CL the $J^P = 0^-$ hypothesis is rejected in favour of the $J^P = 0^+$ hypothesis?

Solution:

$$1. q = -2 \ln \frac{L(H_1)}{L(H_0)}$$

$$3. p_{H_1}(\text{exp}) = 0.37\%$$

$$p_{H_1}(\text{obs}) = 1.5\%$$

$$p_{H_0}(\text{obs}) = 31\%$$

$$4. p_{\mu}^1 = CL_S = \frac{p_{H_1}}{1 - p_{H_0}} = \frac{1.5}{0.69}$$

$$5. CL = 1 - 2.2 = 97.8\%$$