Confidence Interval Estimation
Part II

Roman Kogler (roman.kogler@desy.de)

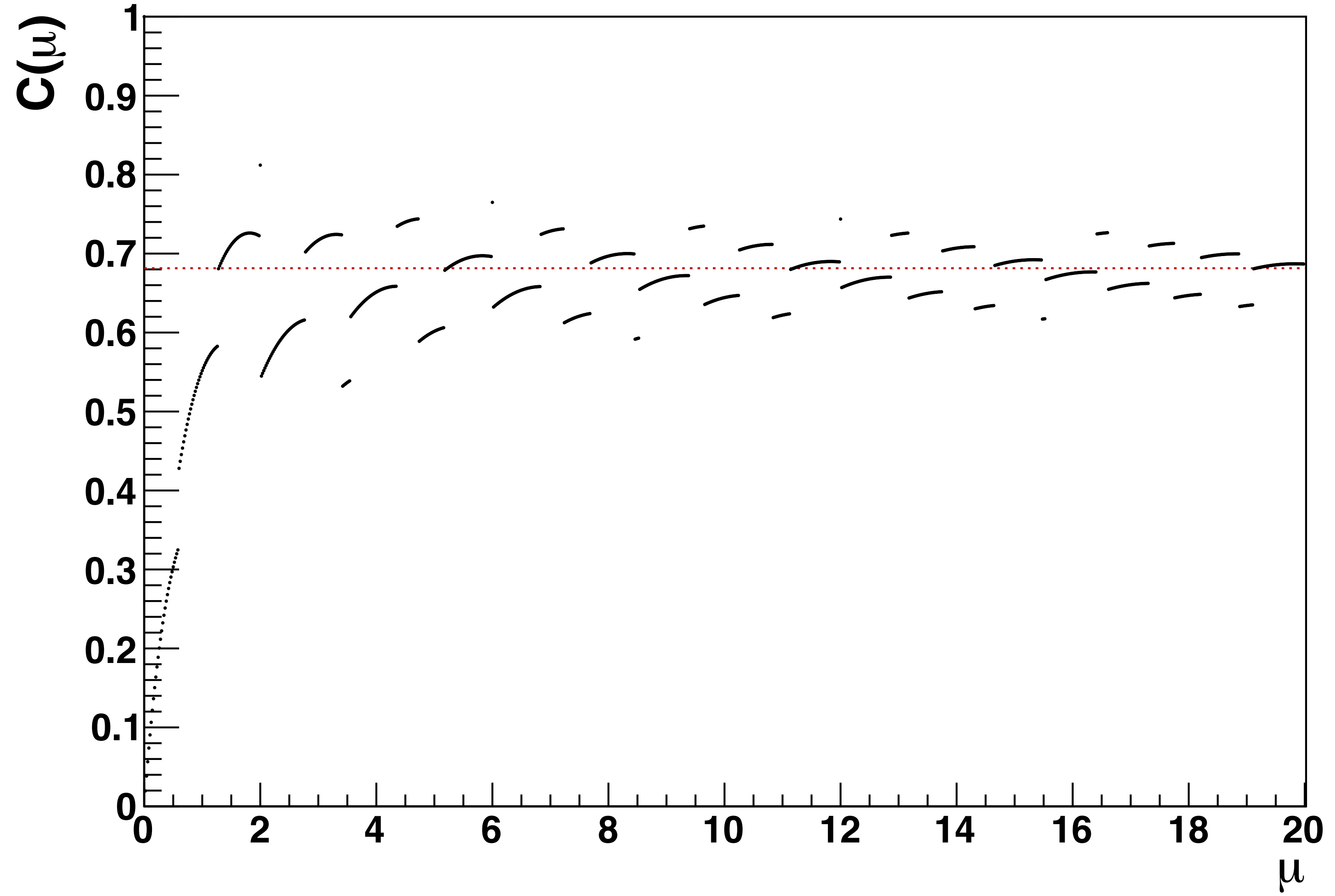
Terascale Statistics School 2023
DESY Hamburg, July 2023

Poisson Distribution: Choice of σ = √n

- Best estimate of μ: Observation n
- Calculate the uncertainty as √n:

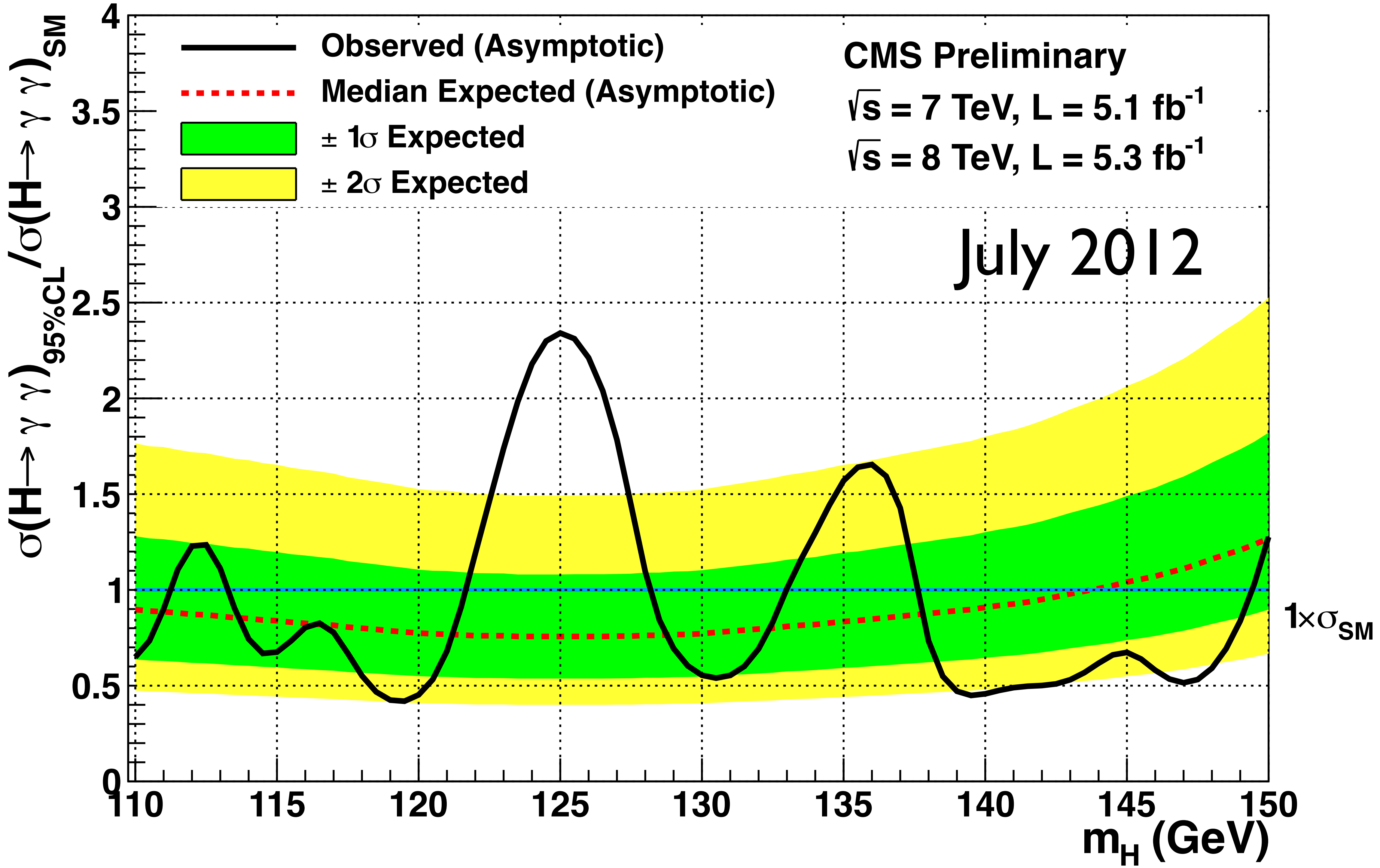
$$n - \sqrt{n} < \mu < n + \sqrt{n}$$

- Serious undercoverage



- minimum coverage is close to 0%

Confidence Intervals and Limits



Note: Upper limits are just one-sided confidence intervals at X% CL

Limits

- Difficulty of setting confidence intervals at physical boundaries (with constraints), for example:

  cross section > 0

  $\theta = x - y$ (for example $m^2 = E^2 - p^2$)

- Assume that we know $\hat{\theta} > 0$, the upper limit is:

  $\theta_{\mathrm{up}} = \hat{\theta} + \sigma_{\hat{\theta}} \Phi^{-1}(1-\beta)$ so for 95% CL $\Phi^{-1}(0.95) = 1.65$

Limits

- Difficulty of setting confidence intervals at physical boundaries (with constraints), for example:

  cross section > 0

  $\theta = x - y$ (for example $m^2 = E^2 - p^2$)

- Assume that we know $\hat{\theta} > 0$, the upper limit is:

  $\theta_{\mathrm{up}} = \hat{\theta} + \sigma_{\hat{\theta}} \Phi^{-1}(1-\beta)$ so for 95% CL $\Phi^{-1}(0.95) = 1.65$

- Now consider the case where we measured $\hat{\theta} = -2$
  and the resolution of our experiment is $\sigma_{\hat{\theta}} = 1$
  We get:

  $\theta_{\mathrm{up}} = -0.35$ at 95% CL and we did not learn anything!

Note that from a frequentist' point of view, this is totally fine. If we performed the experiment a large number of times, we would end up with a meaningful physical bound. This does not change the fact, that we did not learn anything from our experiment at hand.

Possible Solutions

Possible Solutions

1. **Increase the CL**

   fix the negative limit on x by adjusting x_{up} until it becomes positive:

   θ_{up} = 0.326 at 99% CL

Possible Solutions

1. **Increase the CL**

   fix the negative limit on x by adjusting x_{up} until it becomes positive:

   θ_{up} = 0.326 at 99% CL

   $\theta_{up} = 10^{-5}$ at 97.7725% CL

   Results in a solution much better than the experimental resolution! BAD!

Possible Solutions

1. **Increase the CL**

   fix the negative limit on x by adjusting x_{up} until it becomes positive:

   θ_{up} = 0.326 at 99% CL

   θ_{up} = 10^{-5} at 97.7725% CL

   Results in a solution much better than the experimental resolution! BAD!

2. **Shift the measurement**

   $$\theta_{\rm up} = \max\left[\hat{\theta}, 0\right] + \sigma_{\hat{\theta}} \Phi^{-1}(1-\beta)$$

   always results in positive limits on x

Possible Solutions

1. **Increase the CL**

   fix the negative limit on x by adjusting x_{up} until it becomes positive:

   θ_{up} = 0.326 at 99% CL

   θ_{up} = 10^{-5} at 97.7725% CL

   Results in a solution much better than the experimental resolution! BAD!

2. **Shift the measurement**

   $$\theta_{\text{up}} = \max\left[\hat{\theta}, 0\right] + \sigma_{\hat{\theta}}\Phi^{-1}(1-\beta)$$

   always results in positive limits on x

   Coverage probability not given anymore!

Possible Solutions

1. **Increase the CL**

   fix the negative limit on x by adjusting x_{up} until it becomes positive:

   θ_{up} = 0.326 at 99% CL

   θ_{up} = 10^{-5} at 97.7725% CL

   Results in a solution much better than the experimental resolution! BAD!

2. **Shift the measurement**

   $$\theta_{\mathrm{up}} = \max\left[\hat{\theta}, 0\right] + \sigma_{\hat{\theta}} \Phi^{-1}(1-\beta)$$

   always results in positive limits on x

   Coverage probability not given anymore!

3. **Bayes Theorem**

   Can encode our prior knowledge in the prior $\pi(\theta)$!

   $$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{with} \quad P(B) = \sum_i P(B|A_i)P(A_i)$$

Possible Solutions

3. Bayes Theorem

$$p(\theta|\mathbf{d}) = \frac{L(\mathbf{d}|\theta)\pi(\theta)}{\int L(\mathbf{d}|\theta')\pi(\theta')\mathrm{d}\theta'}$$

d… observed data
L… likelihood to observe the data
π… prior knowledge
$p(\theta|\mathbf{d})$… estimator for θ (ML)

Possible Solutions

3. Bayes Theorem

$$p(\theta|\mathbf{d}) = \frac{L(\mathbf{d}|\theta)\pi(\theta)}{\int L(\mathbf{d}|\theta')\pi(\theta')\mathrm{d}\theta'}$$

d… observed data
L… likelihood to observe the data
π… prior knowledge
$p(\theta|\mathbf{d})$… estimator for θ (ML)

Can now calculate an upper limit for given probability α :

$$\alpha = \int_{-\infty}^{\theta_{\mathrm{up}}} p(\theta|\mathbf{d}) \quad \text{and} \quad \beta = \int_{\theta_{\mathrm{down}}}^{\infty} p(\theta|\mathbf{d})$$

and encode the prior knowledge: $\pi(\theta) = \begin{cases} 1 & \theta \geq 0 \\ 0 & \theta < 0 \end{cases}$

Possible Solutions

3. Bayes Theorem

$$p(\theta|\mathbf{d}) = \frac{L(\mathbf{d}|\theta)\pi(\theta)}{\int L(\mathbf{d}|\theta')\pi(\theta')\mathrm{d}\theta'}$$

d… observed data
L… likelihood to observe the data
π… prior knowledge
$p(\theta|\mathbf{d})$… estimator for θ (ML)

Can now calculate an upper limit for given probability α :

$$\alpha = \int_{-\infty}^{\theta_{\mathrm{up}}} p(\theta|\mathbf{d}) \quad \text{and} \quad \beta = \int_{\theta_{\mathrm{down}}}^{\infty} p(\theta|\mathbf{d})$$

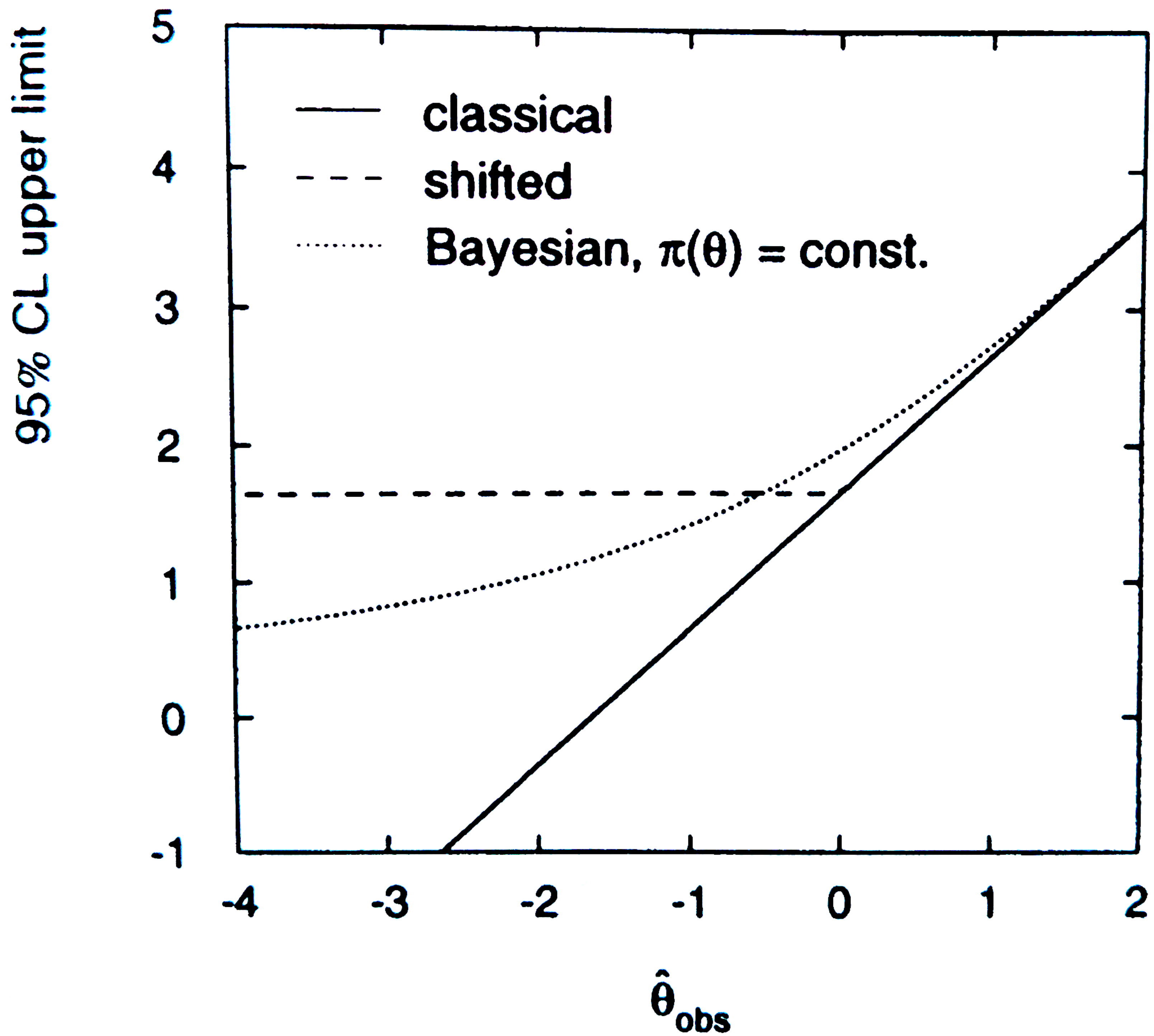
and encode the prior knowledge: $\pi(\theta) = \begin{cases} 1 & \theta \geq 0 \\ 0 & \theta < 0 \end{cases}$

We get:

$$1 - \beta = \frac{1}{N}\int_{-\infty}^{\theta_{\mathrm{up}}} L(\mathbf{d}|\theta')\pi(\theta')\mathrm{d}\theta' \quad \text{with} \quad N = \int_{-\infty}^{\infty} L(\mathbf{d}|\theta')\pi(\theta')\mathrm{d}\theta'$$

solve numerically for θ_{up}

Possible Solutions



picture credit: G. Cowan

Poisson with Background

Realistic case: $n = n_s + n_b$

- Both, n_s and n_b are Poisson variables with means ν_s, ν_b
- Prob. function $f(n;\nu_s,\nu_b) = \frac{(\nu_s+\nu_b)^n}{n!} e^{-(\nu_s+\nu_b)}$

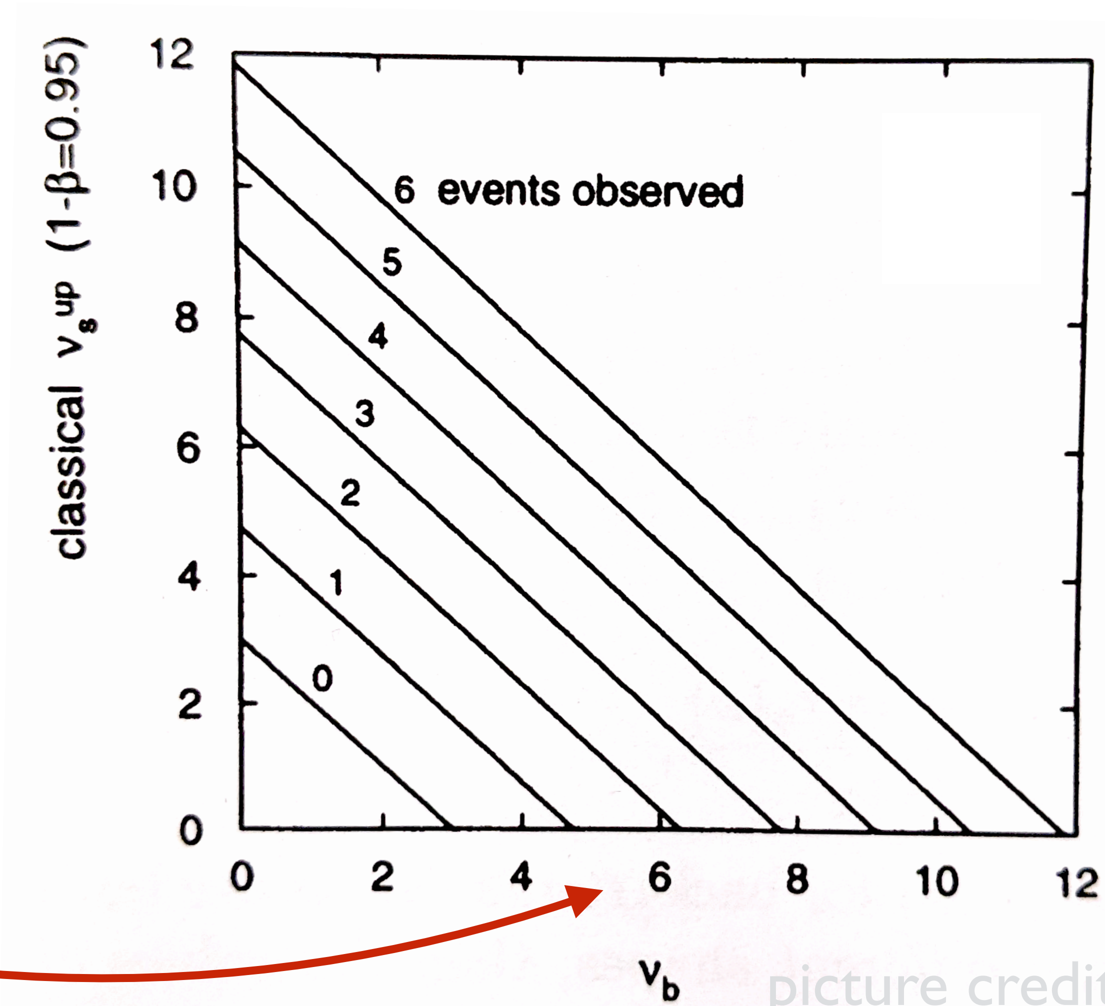
Poisson with Background

Realistic case: $n = n_s + n_b$

- Both, n_s and n_b are Poisson variables with means ν_s, ν_b
- Prob. function $f(n;\nu_s,\nu_b) = \frac{(\nu_s+\nu_b)^n}{n!} e^{-(\nu_s+\nu_b)}$
- Upper limits:

$$\nu_s^{\mathrm{up}} = \nu_s^{\mathrm{up}}(\nu_b = 0) - \nu_b$$

No positive upper limit possible for small n and large ν_b



picture credit: G. Cowan

Poisson with Background

Possible solution: Bayesian method

- Write the likelihood function as a function of ν_s:

$$L(n_{\text{obs}}|\nu_s) = \frac{(\nu_s + \nu_b)^{n_{\text{obs}}}}{n_{\text{obs}}!} e^{-(\nu_s+\nu_b)}$$

- The posterior becomes

$$p(\nu_s|n_{\text{obs}}) = \frac{L(n_{\text{obs}}|\nu_s)\pi(\nu_s)}{\int L(n_{\text{obs}}|\nu_s')\pi(\nu_s')\text{d}\nu_s'}$$

Poisson with Background

Possible solution: Bayesian method

- Write the likelihood function as a function of ν_s:

$$L(n_{\text{obs}}|\nu_s) = \frac{(\nu_s + \nu_b)^{n_{\text{obs}}}}{n_{\text{obs}}!} e^{-(\nu_s+\nu_b)}$$

- The posterior becomes

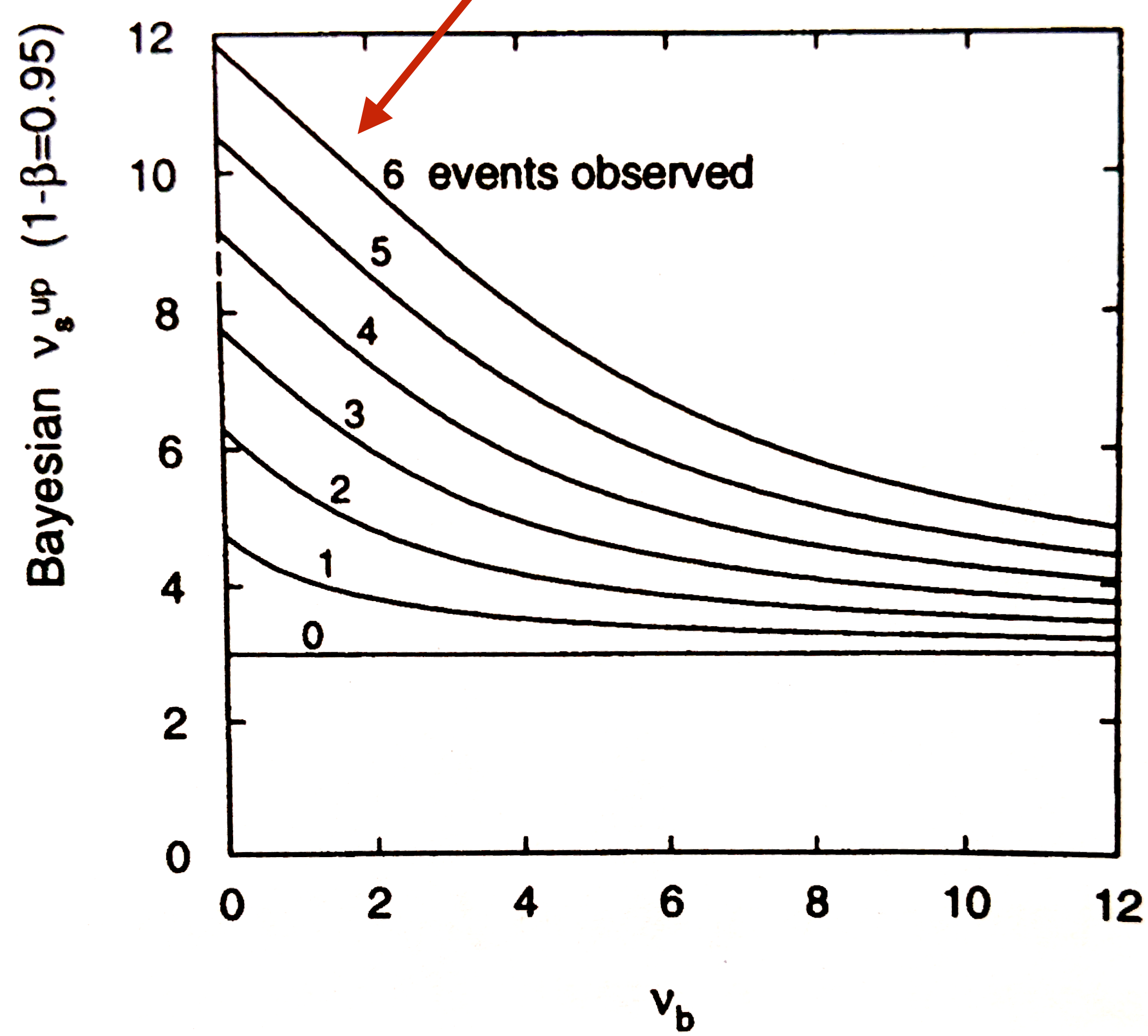
$$p(\nu_s|n_{\text{obs}}) = \frac{L(n_{\text{obs}}|\nu_s)\pi(\nu_s)}{\int L(n_{\text{obs}}|\nu_s')\pi(\nu_s')\mathrm{d}\nu_s'}$$

- Integrate to get an upper limit at a given CL 1-β:

$$\beta = \frac{e^{-(\nu_s^{\text{up}}+\nu_b)} \sum_{n=0}^{n_{\text{obs}}} \frac{1}{n!}(\nu_s^{\text{up}} + \nu_b)^n}{e^{-\nu_b} \sum_{n=0}^{n_{\text{obs}}} \frac{1}{n!}\nu_b^n}$$

(solve numerically for ν_s^{up})

Agreement with previous case (small ν_b)



CLs Method

- Incorporate knowledge about what is signal and background

- Define "test statistic" Q (function of observables and parameters) which "ranks" experiments from the least to most "signal-like"

- Likelihood ratios for signal and background hypotheses are used as the test statistic:

$$\mathcal{Q} = \frac{\mathcal{L}(N_{data}, N_S + N_B)}{\mathcal{L}(N_{data}, N_B)}$$

CLs Method

- Incorporate knowledge about what is signal and background
- Define "test statistic" Q (function of observables and parameters) which "ranks" experiments from the least to most "signal-like"
- Likelihood ratios for signal and background hypotheses are used as the test statistic:

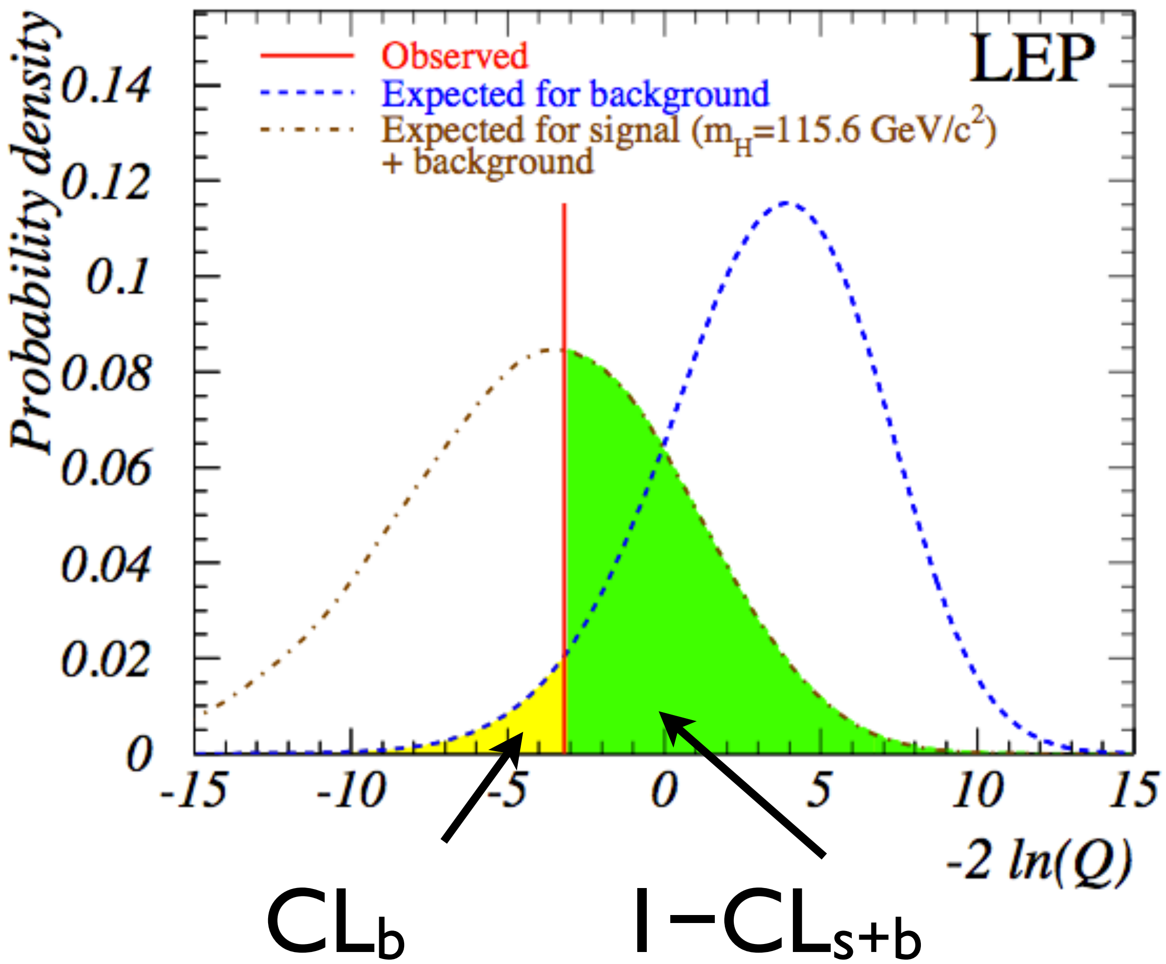
$$\mathcal{Q} = \frac{\mathcal{L}(N_{data}, N_S + N_B)}{\mathcal{L}(N_{data}, N_B)}$$

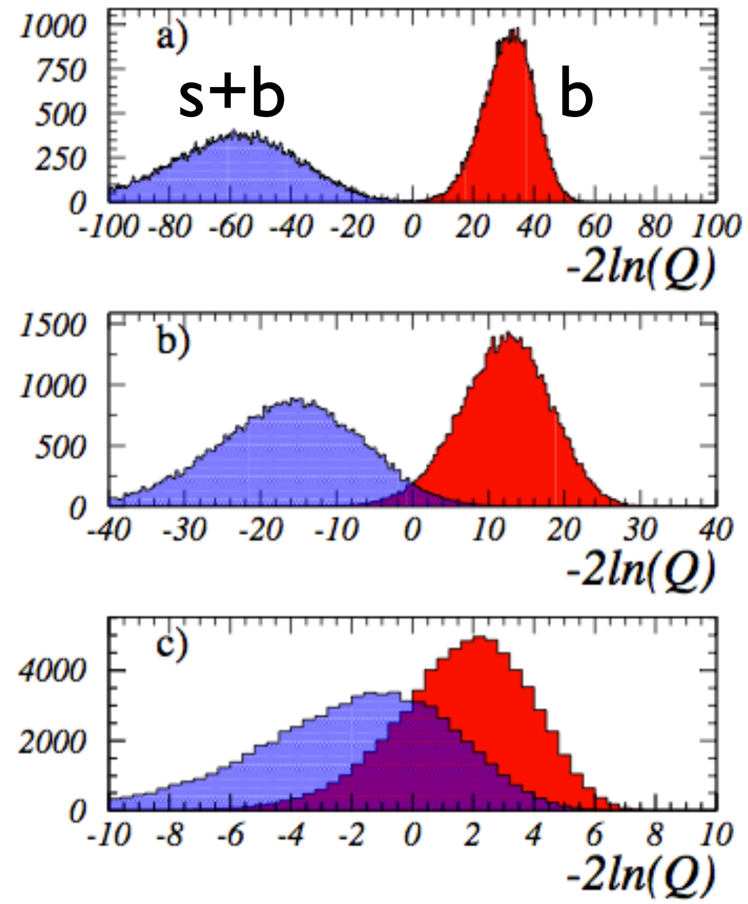
- Confidence in signal+background hypothesis $CL_{s+b} = P_{s+b}(Q \leq Q_{obs})$ where

$$P_{s+b}(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_{s+b}}{dQ} dQ$$

- and dP_{s+b}/dQ is the p.d.f. of Q for the S+B hypothesis
- Confidence in background-only hypothesis: $CL_b = P_b(Q \leq Q_{obs})$
- and analoguos: $P_b(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_b}{dQ} dQ$

CLs Method





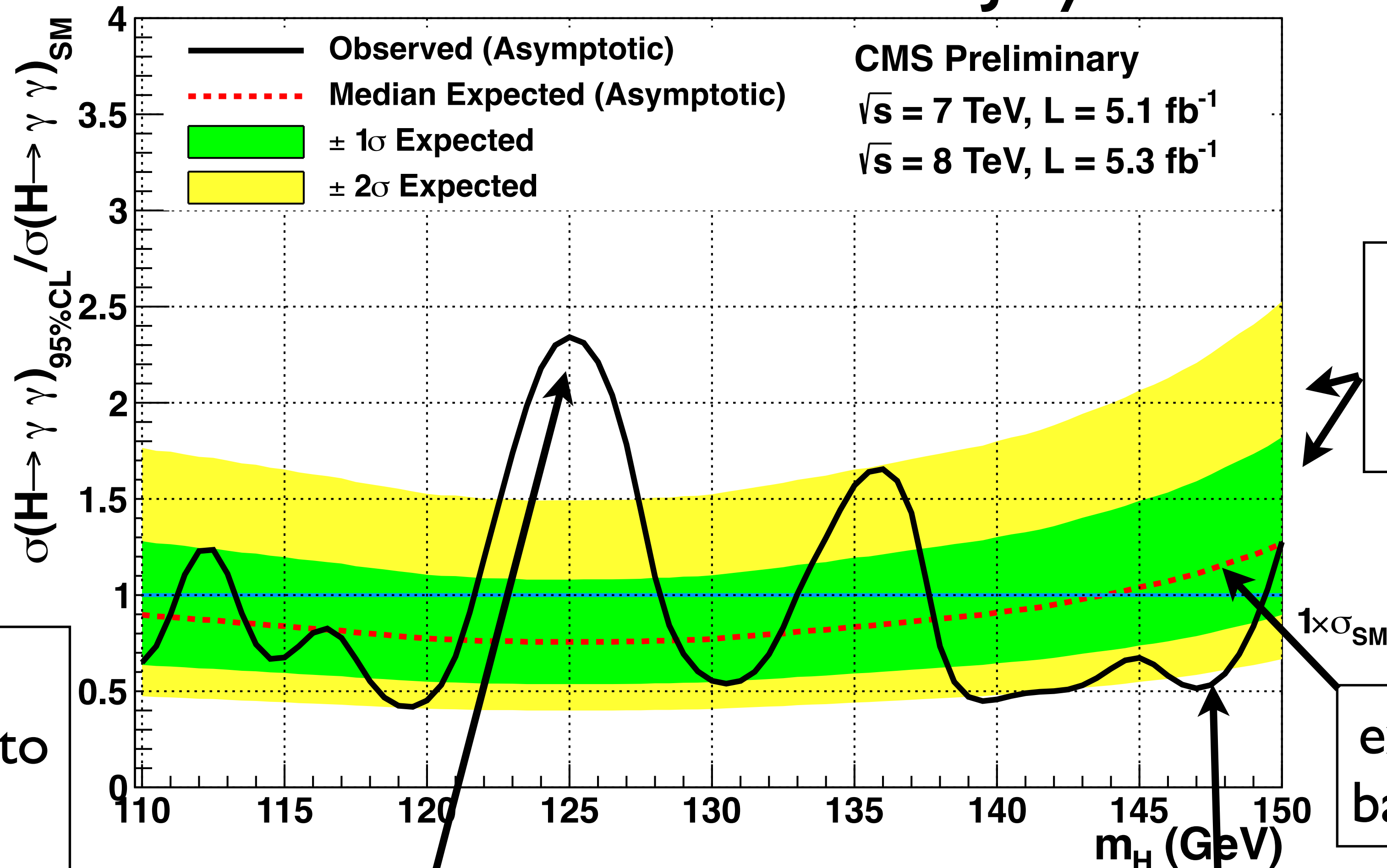
good sensitivity

bad sensitivity

- The problem of unphysical results in case of small signals in the presence of background is avoided by normalising the S+B hypothesis to the B-only hypothesis: $$CL_s = \frac{CL_{s+b}}{CL_b}$$
- Tries to approximate the confidence in the S-hypothesis in the absence of background

Excluding (or Discovering) the Higgs Boson

July 2012



ratio of 95% C.L. excluded x-section to predicted SM x-section

= 1 means that the SM-x-section is excluded (large x-sections are easier to exclude)

here we exclude less than expected (not possible to exclude something that exists!)

here we exclude a mass range even if it is not expected

1 and 2 sigma uncertainties of expected exclusion

expected exclusion based on simulation

this is about exclusion, what about discovery?

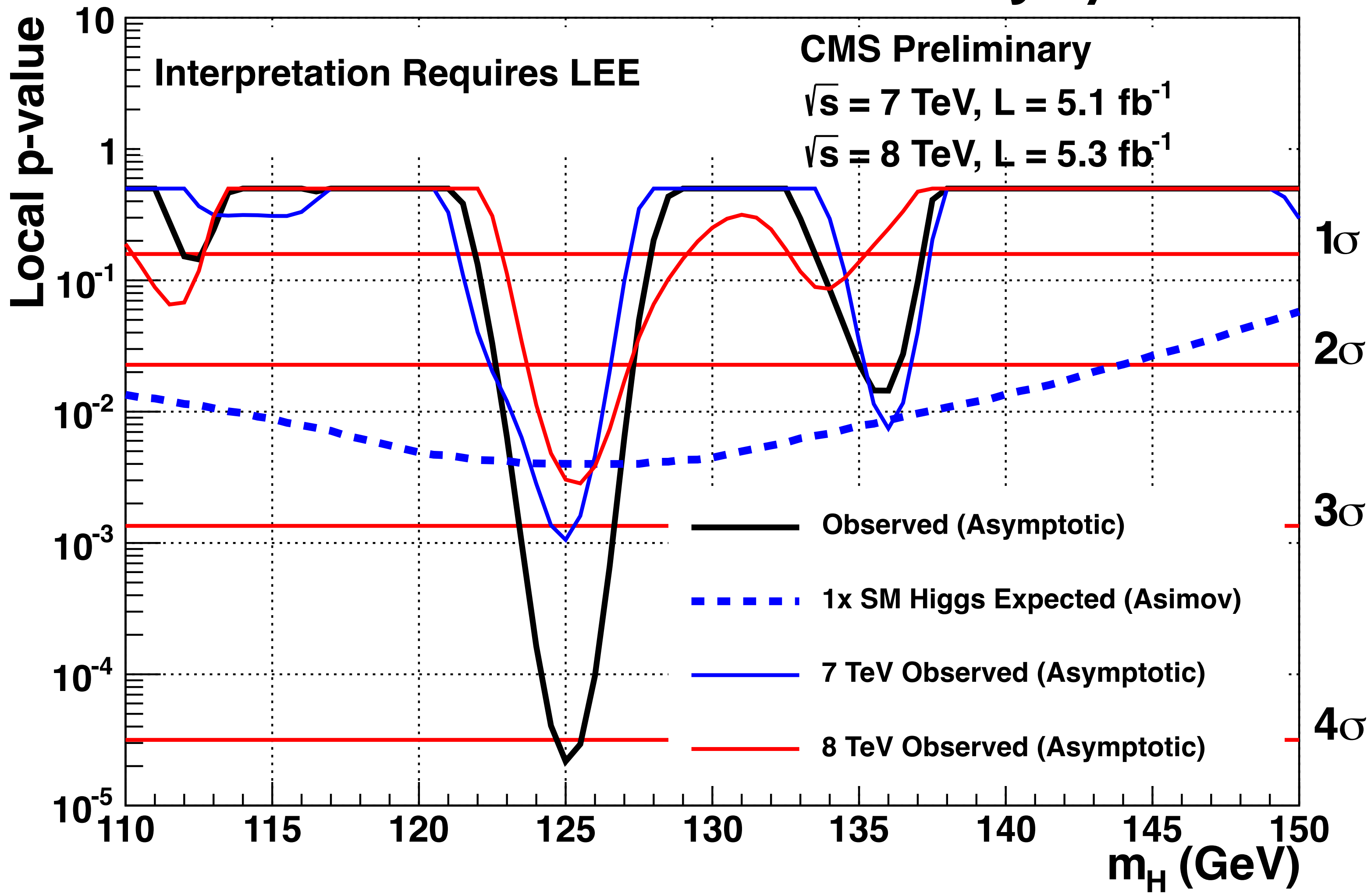
Discovering the Higgs Boson

- Repeat many toy experiments and calculate the probability that the observed excess is a statistical fluctuation

- As a function of the mass

- 3 sigma excesses (called evidence) often disappeared, even though the probability of statistical fluctuation was small….

July 2012



- 5 sigma excesses are called "discovery"