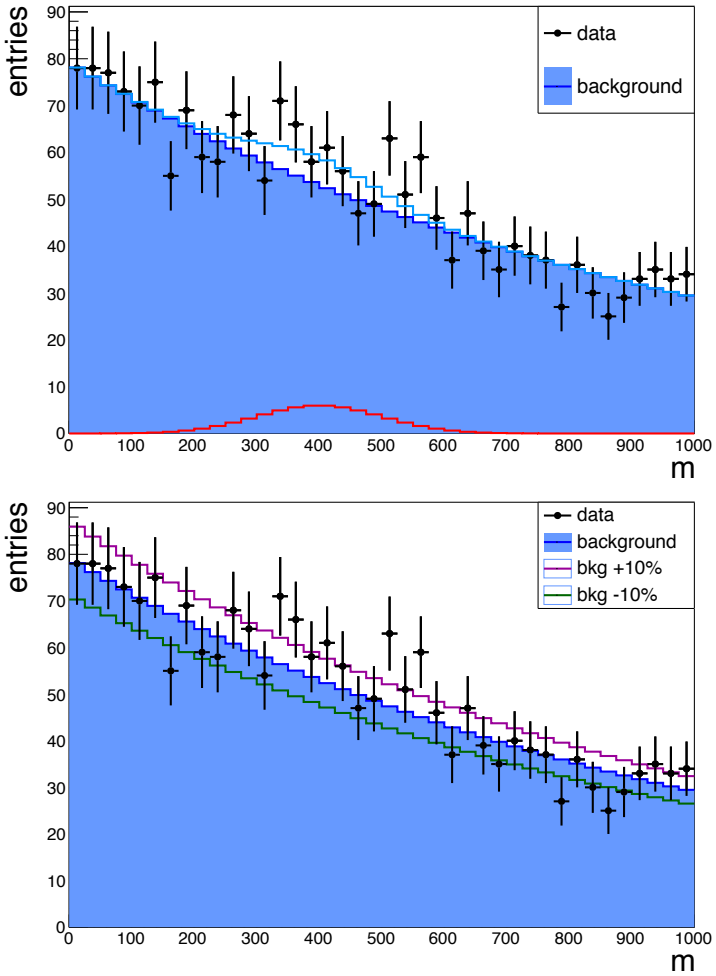
Lecture 3

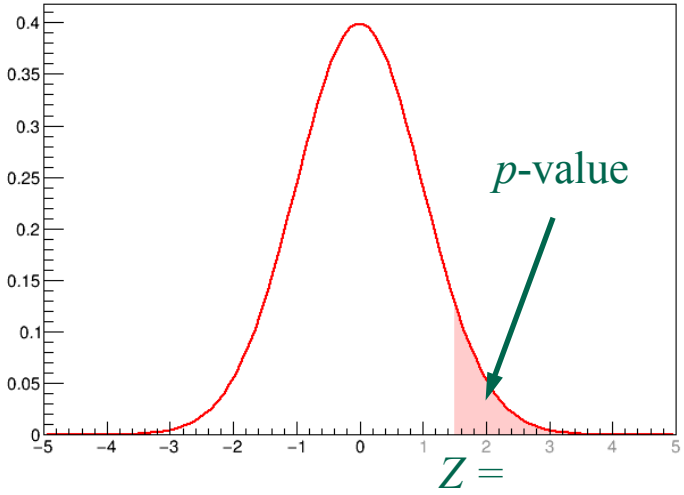
Claiming a discovery

- We want to test our data sample against two hypotheses about the theoretical underlying model:
  - H_0: the data are described by a model that contains background only
  - H_1: the data are described by a model that contains signal plus background
- Our discrimination is based on a test statistic λ whose distribution is known under the two hypotheses
  - Let's assume λ tends to have (conventionally) large values if H_1 is true and small values if H_0 is true
  - This convention is consistent with λ being the likelihood ratio $L(x|H_1)/L(x|H_0)$
- Under the frequentist approach, compute the p-value as the probability that λ is greater or equal to than the value λ_{obs} we observed

Are data below more consistent with a background fluctuation or with a peaking excess?



Significance



- The p-value is usually converted into an equivalent area of a Gaussian tail:

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x = 1 - \Phi(Z)$$

Φ = cumulative of a normal distribution

$$Z = \Phi^{-1}(1-p)$$

- In literature we find, by convention:
  - If the significance is $Z > 3$ ("3σ") one claims "*evidence of*"
    - Probability that background fluctuation will produce a test statistic at least as extreme as the observed value : $p < 1.349 \times 10^{-3}$
  - If the significance is $Z > 5$ ("5σ") one claims "*observation*" (discovery!)
    - $p < 2.87 \times 10^{-7}$
- **Note**: the probability that background produces a large test statistic is not equal to probability of the null hypothesis (background only), which has only a Bayesian sense

Statement by ASA

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

The p-value was never intended to be a substitute for scientific reasoning. Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The ASA statement is intended to steer research into a 'post p < 0.05 era'.

1. *p-values can indicate how incompatible the data are with a specified statistical model.*
2. *p-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.*
3. *Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.*
4. *Proper inference requires full reporting and transparency.*
5. *A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.*
6. *By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.*

Ronald L. Wasserstein, Nicole A. Lazar
The ASA's statement on p-values: context, process, and purpose
DOI:10.1080/00031305.2016.1154108
http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108

Discovery and scientific method

- From Cowan *et al.*, EPJC 71 (2011) 1554:

> *It should be emphasized that in an actual scientific context, rejecting the background-only hypothesis in a statistical sense is only part of discovering a new phenomenon. One's degree of belief that a new process is present will depend in general on other factors as well, such as the plausibility of the new signal hypothesis and the degree to which it can describe the data.*
>
> *Here, however, we only consider the task of determining the p-value of the background-only hypothesis; if it is found below a specified threshold, we regard this as "discovery".*

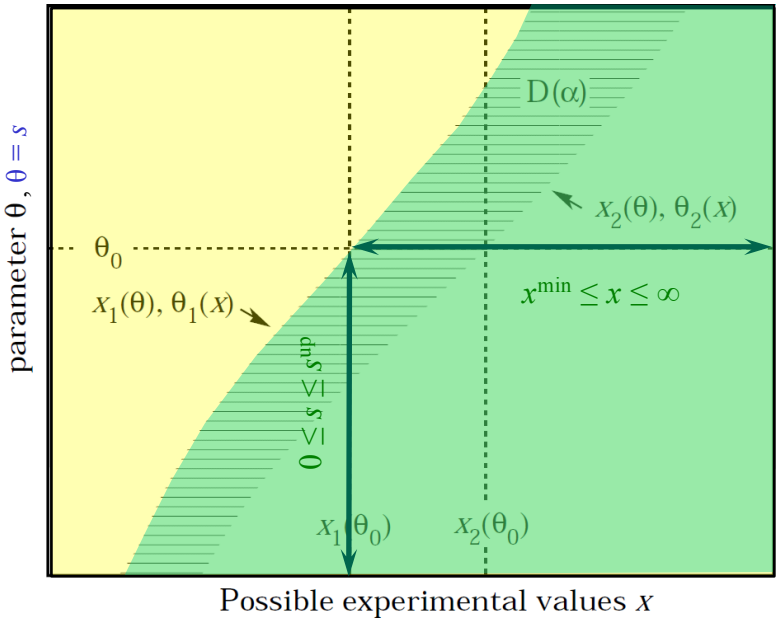
Complementary role of Frequentist and Bayesian approaches ☺

Upper limits

- Measure the amount of excluded region resulting from our (negative) search for a new signal
- Building a fully asymmetric Neyman confidence belt based on the considered test statistic x
- Invert the belt, find the allowed interval:

$$s \in [s_1, s_2] \Rightarrow s \in [0, s^{up}]$$

- Upper limit = upper extreme of the asymmetric interval $[0, s^{up}]$
- In case the observable x is discrete (e.g.: the number of events n in a counting experiments), the coverage may not be exact

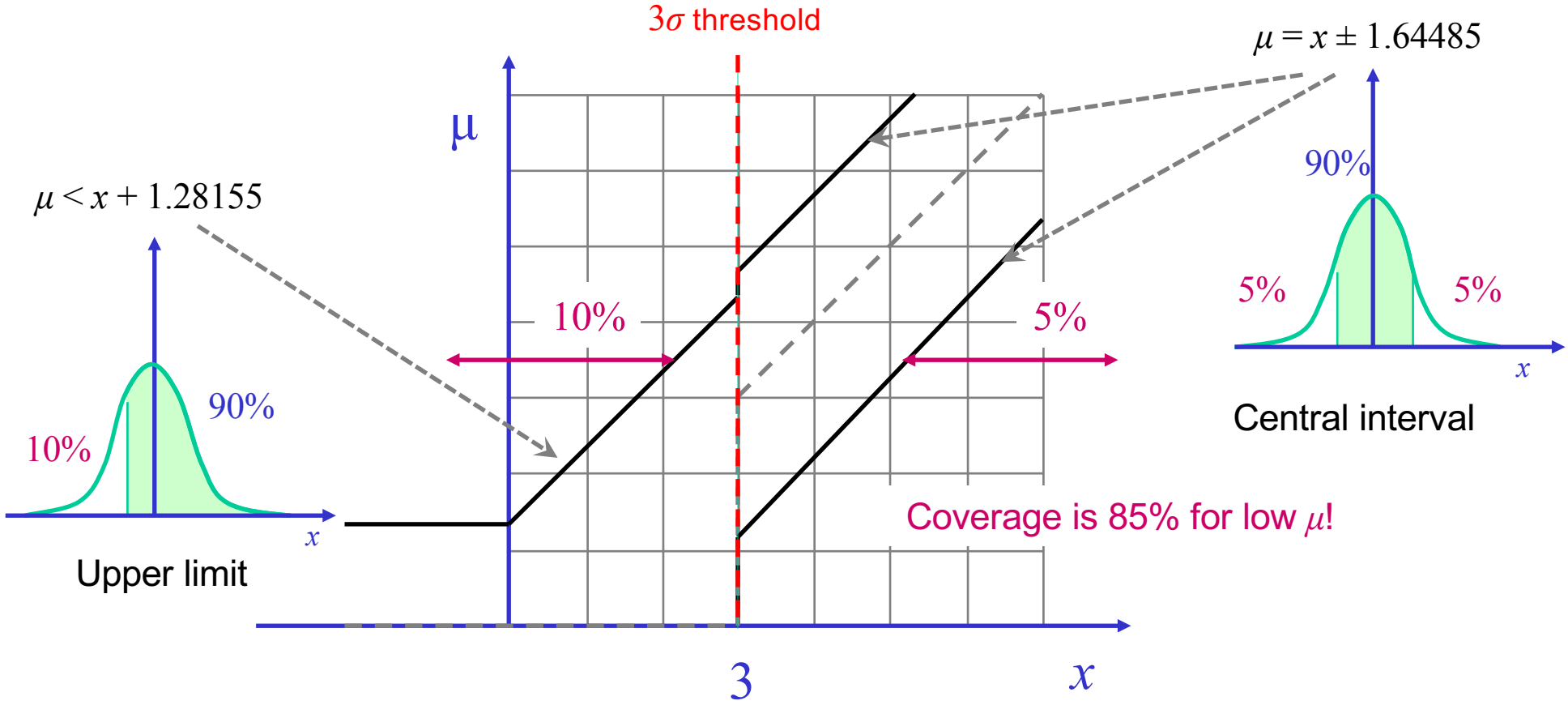


The flip-flopping issue

- When to quote a central value or upper limit?
- A popular choice was:
  - *"Quote a 90% CL upper limit of the measurement if the significance is below 3σ; quote a central value otherwise"*
  - Upper limit ↔ central interval decided according to observed data
- This produces an incorrect coverage!

"Flip-flopping" with a Gaussian PDF

- Assume a Gaussian with a fixed width: $\sigma = 1$



Gary J. Feldman, Robert D. Cousins, Phys.Rev.D57:3873-3889,1998

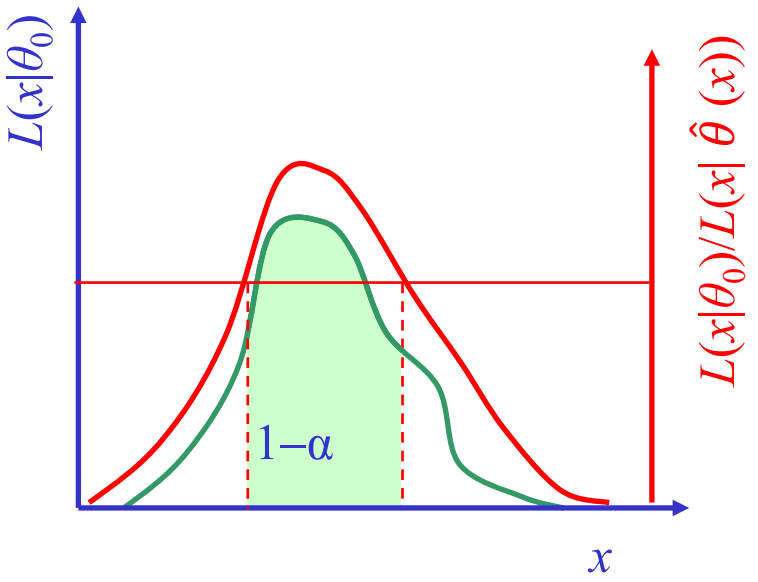
Likelihood ratio & Neyman belt

- Feldman and Cousins proposed a criterion to define the Neyman belt based in a likelihood ratio test:

$$R_\mu = \{x : L(x|\theta_0) \,/\, L(x|\hat{\theta}) > k_\alpha\}$$

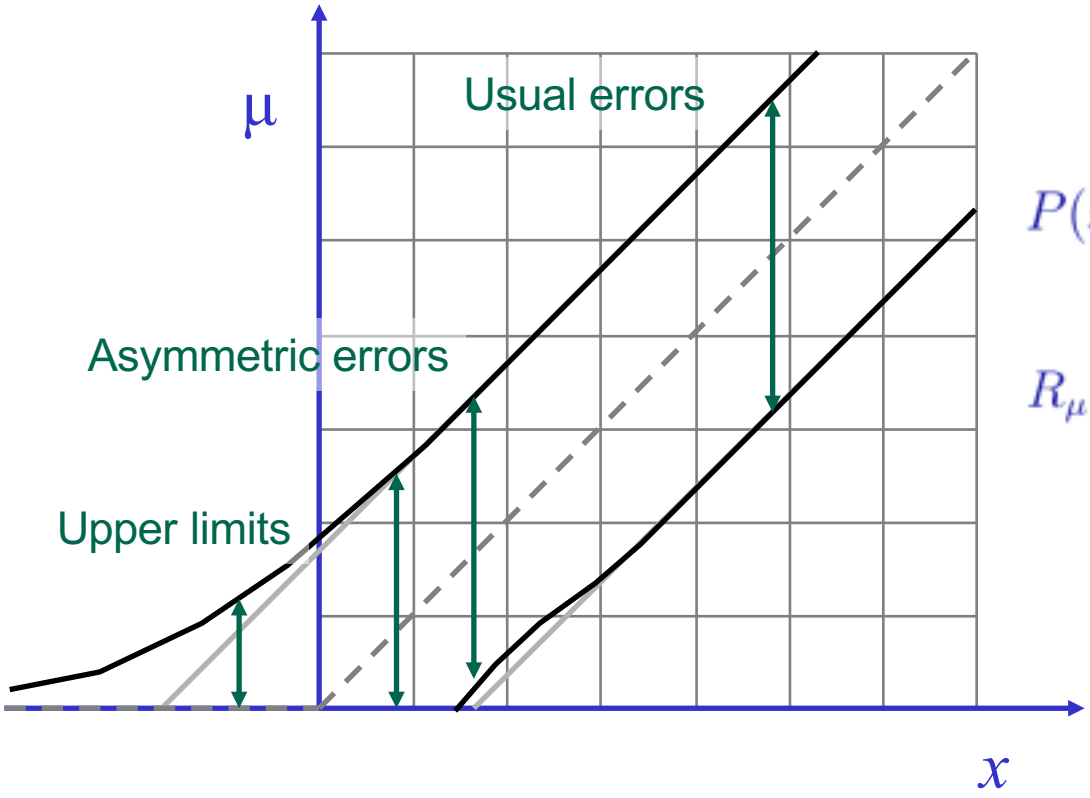
- The value k_α depends on the desired significance level α

- H_0: $\theta = \hat{\theta}$, the best-fit value
- H_1: $\theta = \theta_0$, the specific value considered for the Neyman belt construction



Feldman-Cousins approach

- Application to the Gaussian case:



$$\hat{\mu} = \max(x, 0)$$

$\hat{\mu} = x$ for $x \geq 0$

$$P(x|\hat{\mu}) = \begin{cases} \frac{1}{\sqrt{2\pi}}, & x \geq 0, \\ \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, & x < 0. \end{cases}$$

$$R_\mu(x) = \frac{P(x|\mu)}{P(x|\hat{\mu})} = \begin{cases} e^{-\frac{(x-\mu)^2}{2}}, & x \geq 0, \\ e^{-\frac{x\mu-\mu^2}{2}}, & x < 0. \end{cases}$$

Confidence intervals must be computed numerically, even for this simple Gaussian case!

Upper limits for event counting

- The simplest search for a new signal consists of counting the number of events passing a specified selection
- The number of selected events n is distributed according to a Poissonian distribution
- Expected n for signal + background (H_1): $s + b$
- Expected n for background only (H_0): b

- We measure n events, we want to compare with the two hypotheses H_1 and H_0.
- Simplest case: b is known with negligible uncertainty
  - If not, uncertainty on its estimate must be taken into account

Counting, Bayesian approach

- Let's assume the background b is known with no uncertainty:

$$L(n;s) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

- A uniform prior, $\pi(s) = 1$ simplifies, as usual, the computation:

$$1 - \alpha = \int_0^{s^{\mathrm{up}}} P(s|n)\mathrm{d}s = \frac{\int_0^{s^{\mathrm{up}}} L(n;s)\pi(s)\mathrm{d}s}{\int_0^{\infty} L(n;s)\pi(s)\mathrm{d}s}$$

- Inverting the equation giver the upper limit s^{up}
- For $n = 0$ s^{up} does not depend on b:

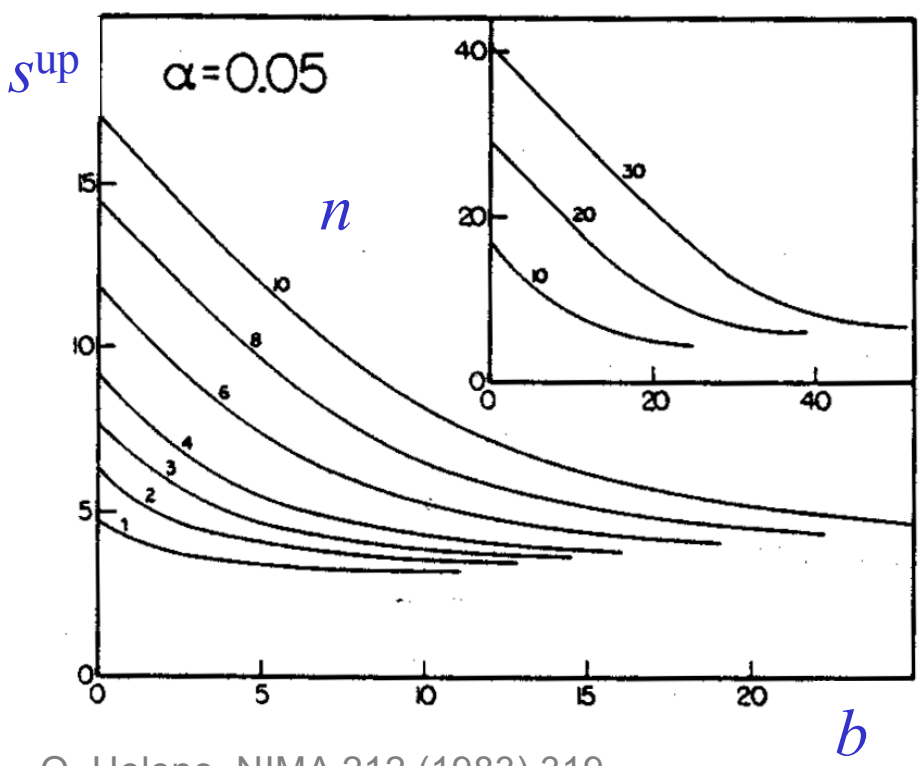
$$\alpha = e^{-s^{\mathrm{up}}}$$

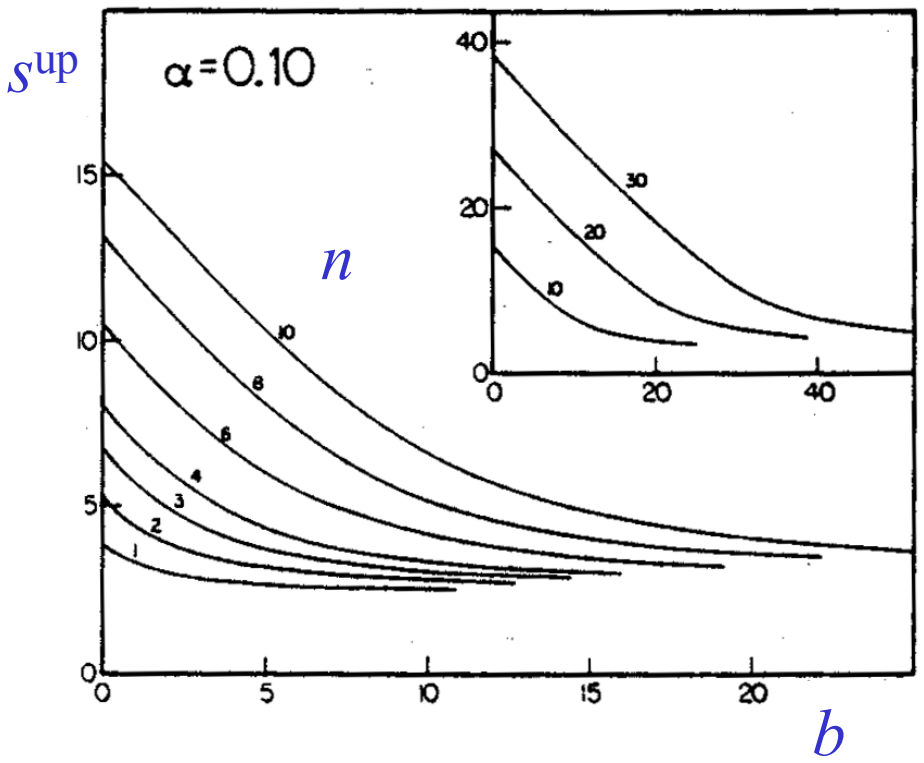
  - $s < 2.303$ (90% CL) ← $\alpha = 0.1$
  - $s < 2.996$ (95% CL) ← $\alpha = 0.05$

$$\alpha = e^{-s^{\mathrm{up}}} \frac{\sum_{m=0}^{n} \frac{(s^{\mathrm{up}}+b)^m}{m!}}{\sum_{m=0}^{n} \frac{b^m}{m!}}$$

Counting, Bayesian approach

- Upper limits decrease as b increases and increase as n increases
- For $n = 0$, upper limits are not sensitive on b (given in prev. slide)





O. Helene. NIMA 212 (1983) 319

Frequentist: zero events selected

- Assume we have negligible background ($b = 0$) and we measure zero events ($n = 0$)
- The likelihood function simplifies as:

$$L(n = 0; s) = \text{Poiss}(0; s) = e^{-s}$$

- The (fully asymmetric) Neyman belt inversion is pretty simple:

$$P(n \leq 0; s^{\text{up}}) = \alpha \rightarrow s^{\text{up}} = -\ln \alpha$$

- The results are by chance identical to the Bayesian computation:

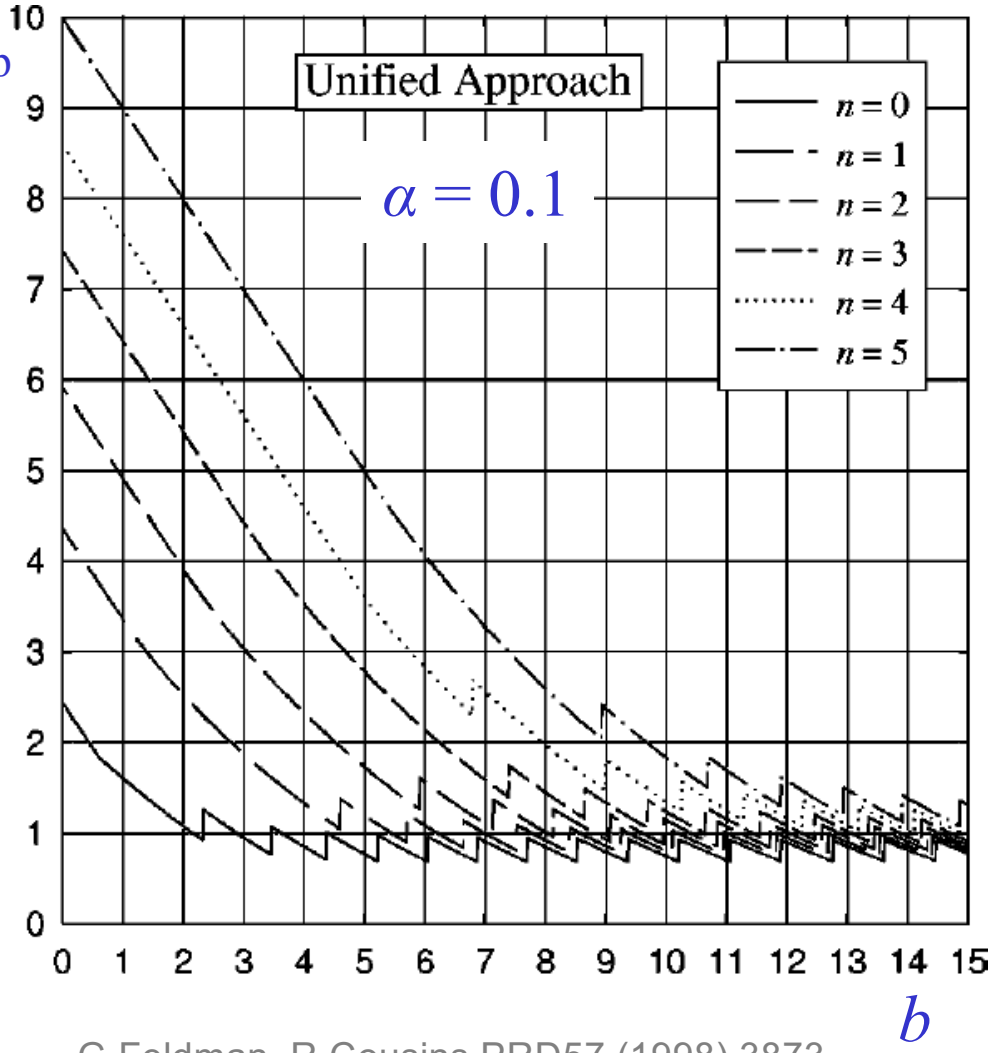
$$s < 2.303 \ (90\% \ \text{CL}) \leftarrow \alpha = 0.1$$

$$s < 2.996 \ (95\% \ \text{CL}) \leftarrow \alpha = 0.05$$

- In spite of the numerical coincidence, the interpretation of frequentist and Bayesian upper limits remain very different!
- **Warning**: this evaluation suffer from the "**flip-flopping**" problem, so the coverage is spoiled if you decide to switch from upper limit to a central value depending on the observed significance!

Conting, Feldman-Cousins

- F&C intervals cure the flip-flopping issue and ensure the correct coverage
  - May overcover for discrete variables
- The "ripple" structure is due to the discrete nature of Poissonian counting
- Note that even for $n = 0$ the upper limit decrease as b increases (apart from ripple effects)
- If two experiment are designed for an expected background of –say– 0.5 and 0.01, the "worse" one has the best expected upper limit



G.Feldman, R.Cousins PRD57 (1998) 3873
C. Giunti, PRD59 (1999), 053001

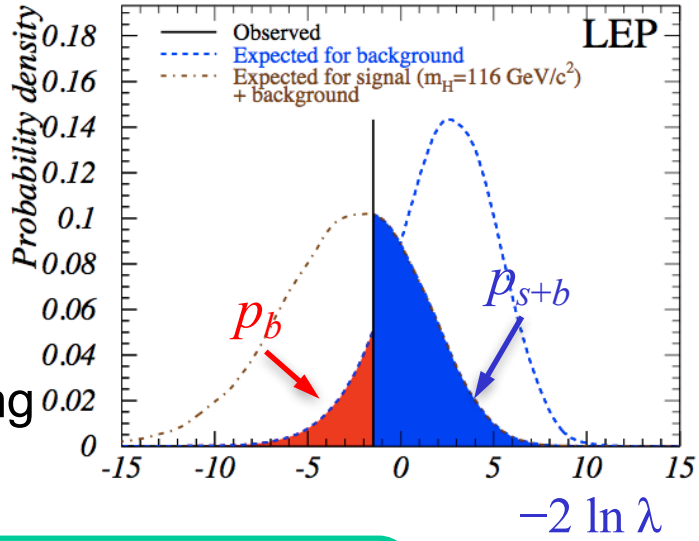
From PDG Review…

- *"The intervals constructed according to the unified procedure [FC] for a Poisson variable n consisting of signal and background have the property that for n = 0 observed events, the upper limit decreases for increasing expected background. This is counter-intuitive, since it is known that if n = 0 for the experiment in question, then no background was observed, and therefore one may argue that the expected background should not be relevant. The extent to which one should regard this feature as a drawback is a subject of some controversy"*

Modified frequentist approach

- A modified approach was proposed for the first time when combining the limits on the Higgs boson search from the four LEP experiments, ALEPH, DELPHI, L3 and OPAL
- Given a test statistic $\lambda(x)$, determine its distribution for the two hypotheses $H_1(s+b)$ and $H_0(b)$, and compute:

$$\begin{cases} p_{s+b} = P\left(\lambda(x|H_1) \leq \lambda^{\mathrm{obs}}\right) \\ p_b = P(\lambda(x|H_0) \geq \lambda^{\mathrm{obs}}) \end{cases}$$



- The upper limit is computed, instead of requiring $p_{s+b} \leq \alpha$, on the modified statistic $\mathrm{CL}_s \leq \alpha$:

$$\mathrm{CL}_s = \frac{p_{s+b}}{1 - p_b}$$

- Since $1 - p_b \leq 1$, $\mathrm{CL}_s \geq p_{s+b}$, hence upper limits computed with the CL_s method are always **conservative**

Note: $\lambda \leq \lambda^{\mathrm{obs}}$ implies $-2\ln\lambda \geq \lambda^{\mathrm{obs}}$

CL_s with toy experiments

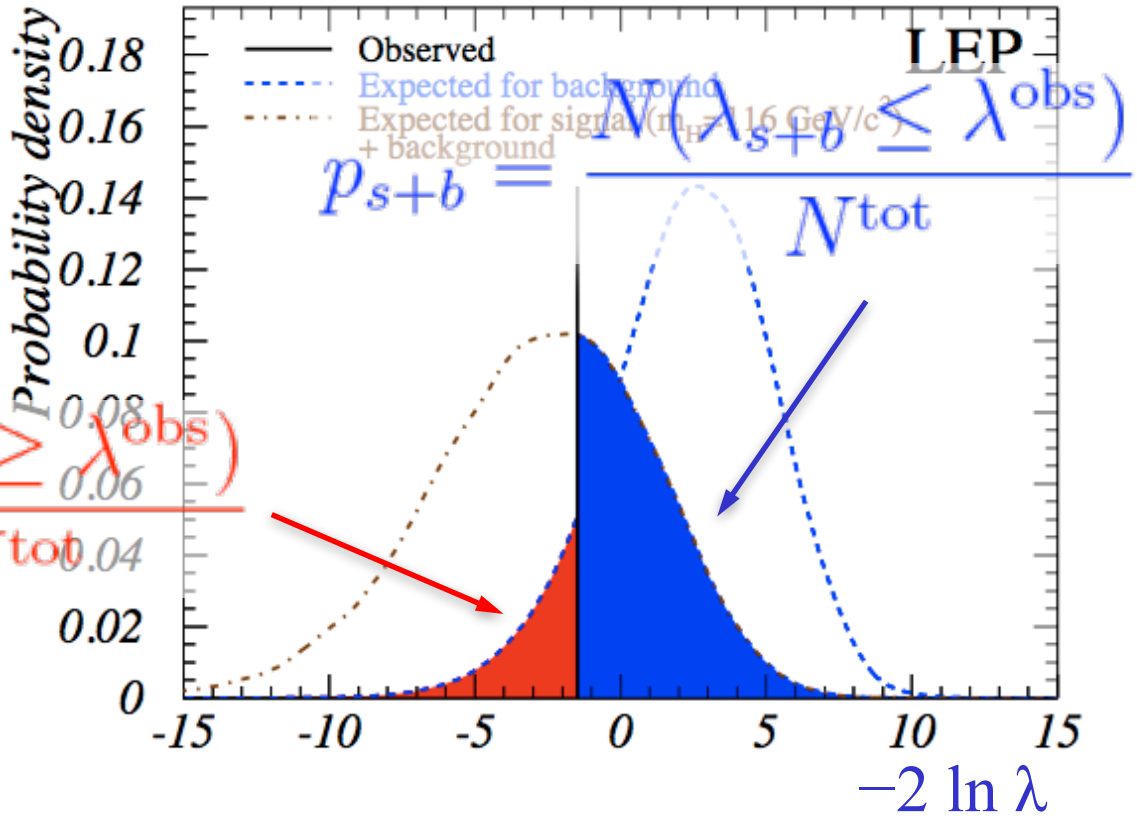
- In practice, p_b and p_{s+b} are computed in from simulated pseudo-experiments ("toy Monte Carlo")

$$CL_s = \frac{N(\lambda_{s+b} \leq \lambda^{\rm obs})}{N(\lambda_b \leq \lambda^{\rm obs})}$$

$$p_b = \frac{N(\lambda_b \geq \lambda^{\rm obs})}{N^{\rm tot}}$$

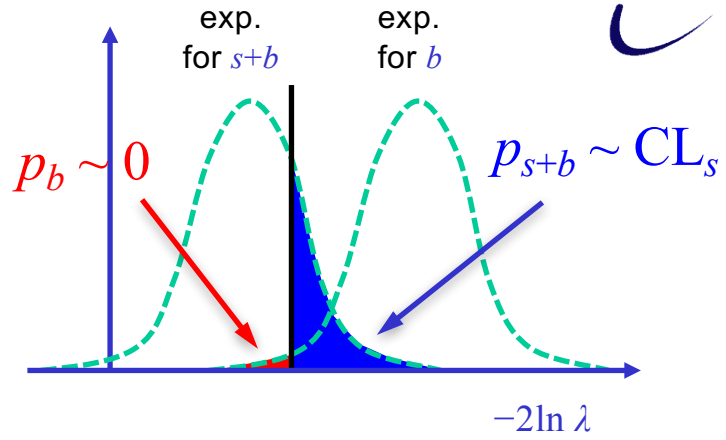
$$p_{s+b} = \frac{N(\lambda_{s+b} \leq \lambda^{\rm obs})}{N^{\rm tot}}$$

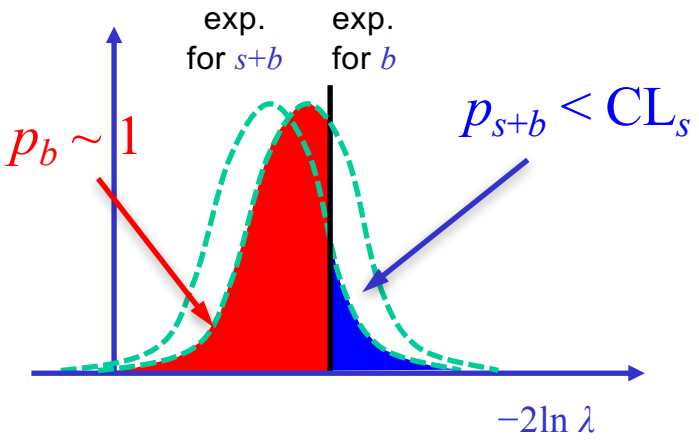
Plot from LEP Higgs combination paper



Main CL_s features

- p_{s+b}: probability to obtain a result which is less compatible with the signal than the observed result, assuming the signal hypothesis
- p_b: probability to obtain a result less compatible with the background-only hypothesis than the observed one
- If the two distributions are very well separated ad H_1 is true, than p_b will be very small $\Rightarrow$ $1-p_b \sim 1$ and $CL_s \sim p_{s+b}$, i.e: the ordinary p-value of the $s+b$ hypothesis
- If the two distributions largely overlap, than if p_b will be large $\Rightarrow$ $1 - p_b$ small, preventing CL_s to become very small





- $CL_s < 1 - \alpha$ prevents rejecting cases where the experiment has little sensitivity

$$CL_s = \frac{p_{s+b}}{1-p_b} = \frac{P(\lambda_{s+b} \leq \lambda^{\text{obs}})}{P(\lambda_b \leq \lambda^{\text{obs}})}$$

Event counting with CL_s

- Let's consider the previous event counting experiment, using $n = n^{obs}$ as test statistic
- In this case CL_s can be written as:

$$CL_s = \frac{P(n \leq n^{obs}|s+b)}{P(n \leq n^{obs}|b)}$$

- Explicitating the Poisson distribution, the computation gives the same result as for the Bayesian case with a uniform prior
- In many cases the CL_s upper limits give results that are very close, numerically, to Bayesian computations done assuming a uniform prior
- But the interpretation is very different from Bayesian limits!

$$\alpha = e^{-s^{up}} \frac{\sum_{m=0}^{n} \frac{(s^{up}+b)^m}{m!}}{\sum_{m=0}^{n} \frac{b^m}{m!}}$$

Observations on the CL_s method

- *"A specific modification of a purely classical statistical analysis is used to avoid excluding or discovering signals which the search is in fact not sensitive to"*
- *"The use of CLs is a conscious decision not to insist on the frequentist concept of full coverage (to guarantee that the confidence interval doesn't include the true value of the parameter in a fixed fraction of experiments)."*
- *"confidence intervals obtained in this manner do not have the same interpretation as traditional frequentist confidence intervals nor as Bayesian credible intervals"*

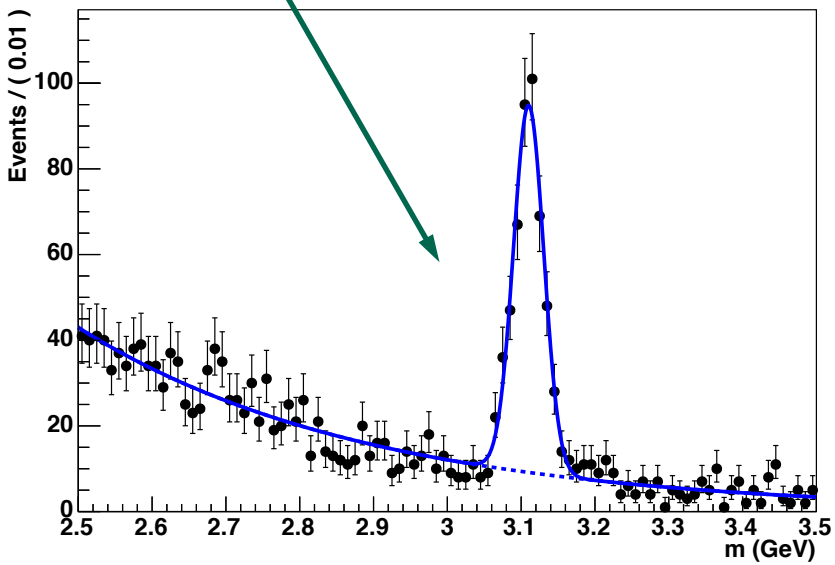
A. L. Read, Modified frequentist analysis of search results (the CLls method), 1st Workshop on Confidence Limits, CERN, 2000

Nuisance parameters

- Usually, signal extraction procedures (fits, upper limits setting) determine, together with parameters of interest, also nuisance parameters that model effects not strictly related to our final measurement
  - Background yield and shape parameters
  - Detector resolution
  - ...
- Nuisance parameters are also used to model sources of systematic uncertainties
- Often referred to nominal values
  - Examples:
  - $b = \beta\, \sigma_b\, L_{\rm int}$ with $\beta^{\rm nominal} = 1$
  - $b = e^{\beta}\, \sigma_b\, L_{\rm int}$ with $\beta^{\rm nominal} = 0$ (negative yields not allowed!)

cross section × int. lumi

$$L(m; s, b, \mu, \sigma, \lambda) = \frac{e^{-(s+b)}}{n!}\left(s\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(m-\mu)^2}{2\sigma^2}} + b\lambda e^{-\lambda m}\right)$$



Nuisance pars in Bayesian approach

- Notation below: μ = parameter(s) of interest, θ = nuisance parameter(s)
- No special treatment:

$$P(\mu,\theta|x) = \frac{L(x;\mu,\theta)\pi(\mu,\theta)}{\int L(x;\mu',\theta')\pi(\mu',\theta')\mathrm{d}\mu'\mathrm{d}\theta'}$$

- $P(\mu|x)$ obtained as marginal PDF of μ obtained integrating on θ:

$$P(\mu|x) = \int P(\mu,\theta|x)\mathrm{d}\theta = \frac{\int L(x;\mu,\theta)\pi(\mu,\theta)\mathrm{d}\theta}{\int L(x;\mu',\theta)\pi(\mu',\theta)\mathrm{d}\mu'\mathrm{d}\theta}$$

Nuisance pars., frequentist

- Introduce a complementary dataset to constrain the nuisance parameters θ (e.g.: calibration data, background estimates from control sample…)
- Formulate the statistical problem in terms of both the main data sample (x) and the control sample (y)

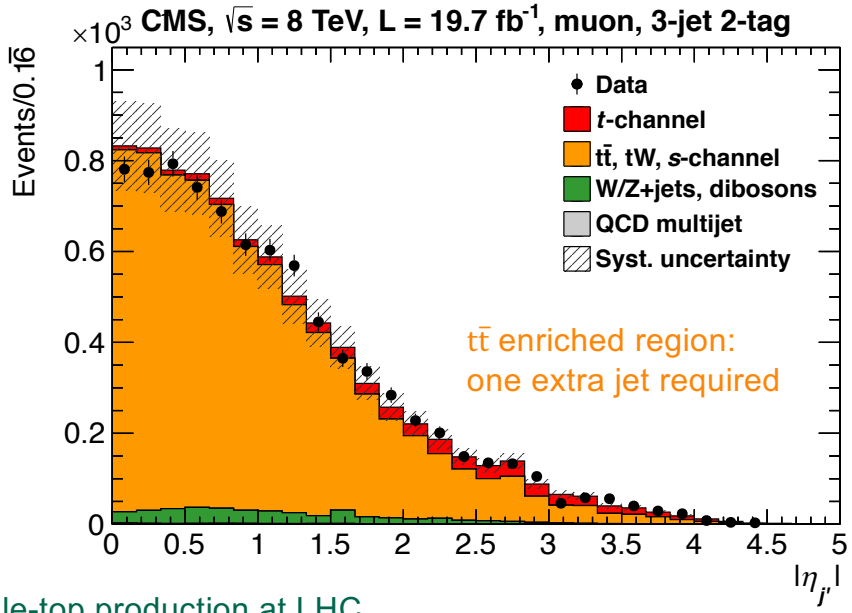
$$L(x, y; \mu, \theta) = L(x; \mu, \theta) L(y; \theta)$$

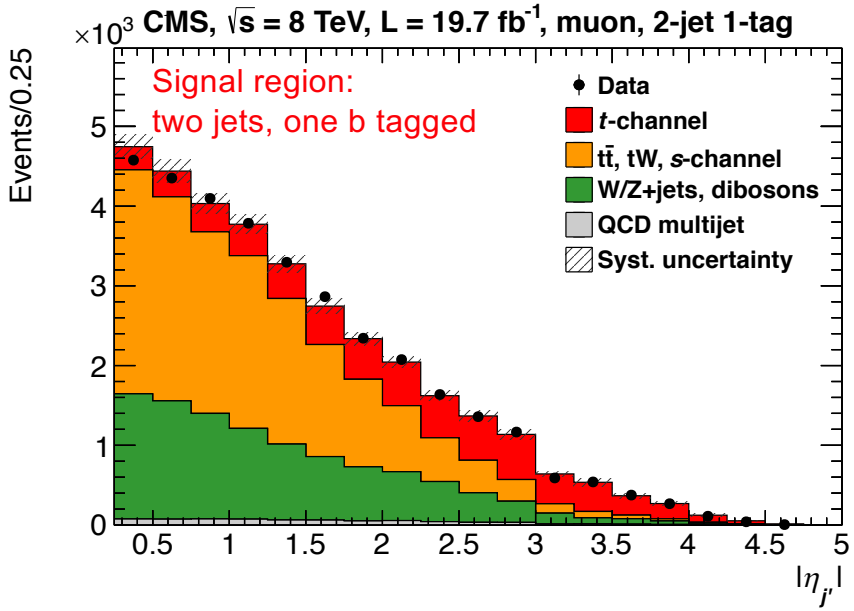
- Not always the control sample data are available
  - E.g.: calibration from test beam, stored in different formats, control samples analyzed with different software framework…
  - In some cases may be complex and CPU intensive
- Simplest case; assume known PDF for "nominal" value of θ^{nom} (e.g.: estimate with Gaussian uncertainty)

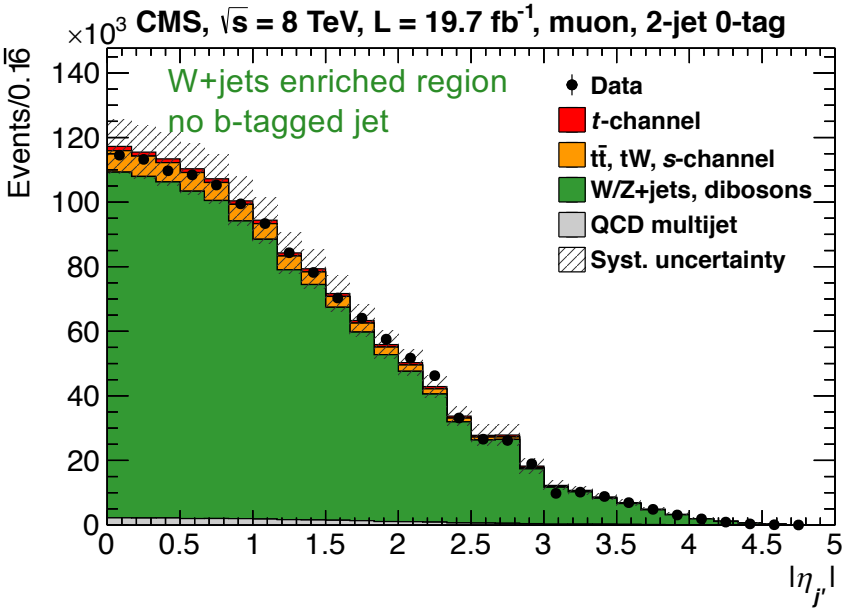
$$L(x, \theta^{\mathrm{nom}}; \mu, \theta) = L(x; \mu, \theta) L(\theta^{\mathrm{nom}}; \theta)$$

Fitting control regions

- In some cases, background parameters can be constrained from statistically independent control samples
  - Consider possible signal contamination!
- Background yield can be measured in background-enriched regions and extrapolated to signal regions applying scale factors predicted by simulation
- Complete likelihood function = product of likelihood functions in each considered regions, sharing common nuisance parameters
  - Typically: background rates







Cousins-Highland hybrid approach

- Method proposed by Cousins and Highland
  - Add posterior from another experiment into the likelihood definition
  - Integrate the likelihood function over the nuisance parameters

$$L^{\text{hybrid}}(x;\mu) = \int L(x;\mu,\theta) L(\theta^{\text{nom}};\theta)\mathrm{d}\theta$$

- Also called "hybrid" approach, because a partial Bayesian approach is implicit in the integration
  - Bayesian integration of PDF, then likelihood used in a frequentist way
- Not guaranteed to provide exact frequentist coverage!
- Numerical studies with pseudo experiments showed that the hybrid CL_s upper limits gives very similar results to Bayesian limit assuming a uniform prior

NIM A320 (1992) 331-335

Profile likelihood

- Define a test statistic based on a likelihood ratio:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

$L(\mu, \hat{\hat{\theta}})$ ← Fix μ, fit θ

$L(\hat{\mu}, \hat{\theta})$ ← Fit both μ and θ

- μ is usually the "signal strength" (i.e.: $\sigma/\sigma_{\text{th}}$) in case of a search for a new signal
- Different 'flavors' of test statistics
  - E.g.: deal with unphysical $\mu < 0$, …
- The distribution of $q_\mu = -2 \ln \lambda(\mu)$ may be asymptotically approximated to the distribution of a χ^2 with one degree of freedom (one parameter of interest = μ) due to the Wilks' theorem

(→ next slide)

Wilks' theorem (1938)

- Consider a likelihood function from N measurements:

$$\prod_{i=1}^{N} L(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m) = \prod_{i=1}^{N} L(\vec{x}_i; \vec{\theta})$$

- Assume that H_0 and H_1 are two nested hypotheses, i.e.: they can be expressed as:

$$\vec{\theta} \in \Theta_0 \qquad \vec{\theta} \in \Theta_1$$

- Where $\Theta_0 \subseteq \Theta_1$. Then, the following quantity for $N \to \infty$ is distributed as a χ^2 with n.d.o.f. equal to the difference of Θ_0 and Θ_1 dimensionality:

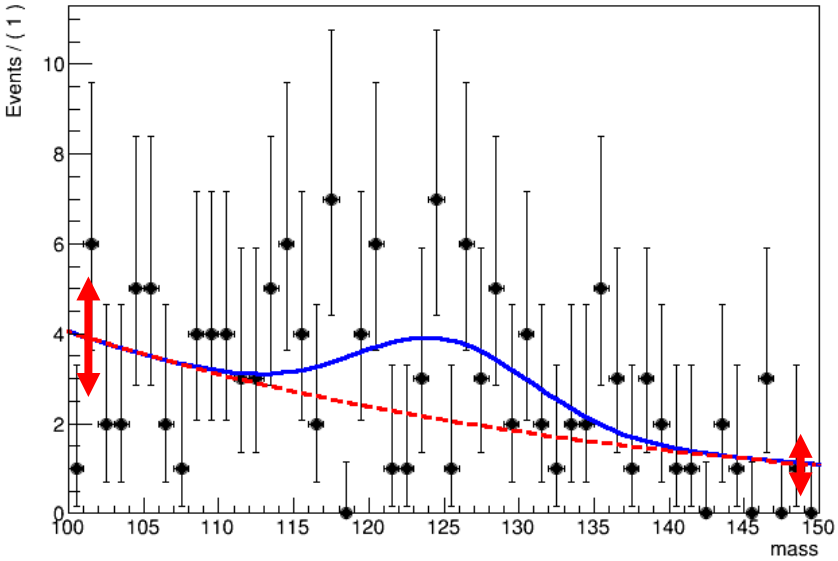
$$\chi_r^2 = -2 \ln \frac{\sup_{\vec{\theta} \in \Theta_0} \prod_{i=1}^{N} L(\vec{x}_i; \vec{\theta})}{\sup_{\vec{\theta} \in \Theta_1} \prod_{i=1}^{N} L(\vec{x}_i; \vec{\theta})}$$

- E.g.: searching for a signal with strength μ, $H_0: \mu = 0$, $H_1: \mu \geq 0$ we have the profile likelihood (supremum = best fit value):

$$\chi_r^2(\mu) = -2 \ln \frac{\sup_{\vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu, \vec{\theta})}{\sup_{\mu', \vec{\theta}} \prod_{i=1}^{N} L(\vec{x}_i; \mu', \vec{\theta})}$$

Systematic uncertainties

- Gaussian signal over an exponential background
- Fix all parameters from theory prediction, fit only the signal yield
- Assume a –say– 30% uncertainty on the background yield
- A log normal model may be assumed to avoid unphysical negative yields
  - $b_0 = b\, e^{\beta}$, where our estimate β is known with a Gaussian uncertainty $\sigma_\beta = 0.3$



b_0 = true (unknown) value

b = our estimate

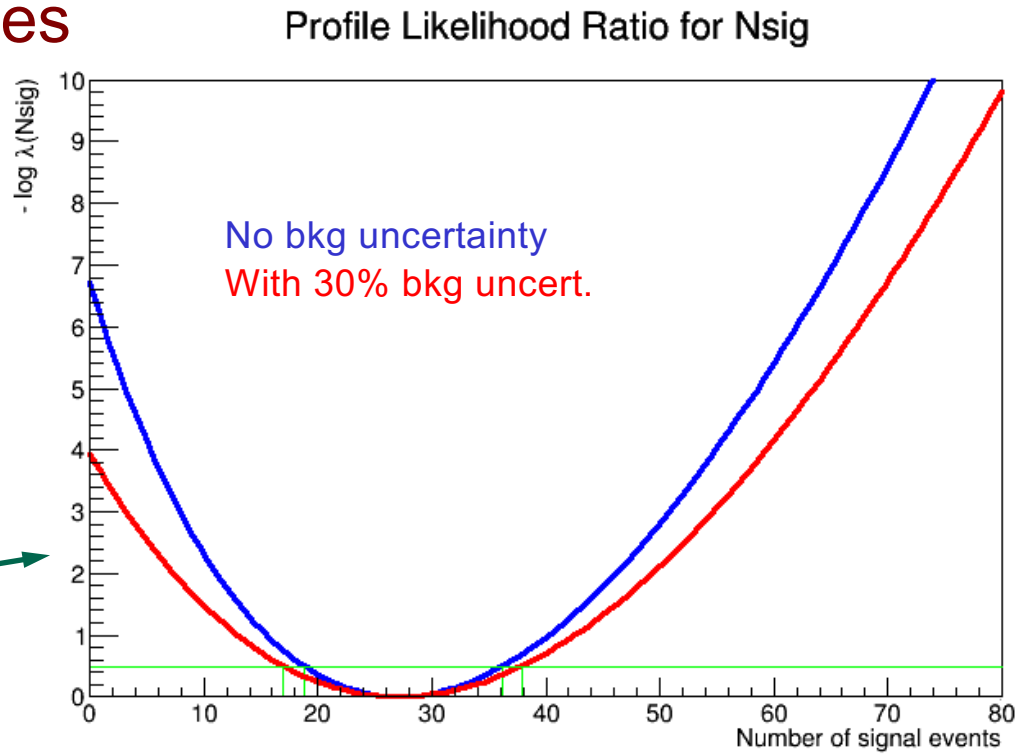
$$L(m; s, \beta) = L_0(m; s, b_0 = be^{\beta}) P(\beta; \sigma_\beta)$$

$$L_0(m; s, b_0) = \frac{e^{-(s+b_0)}}{n!} \left(s \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(m-\mu)^2}{2\sigma^2}} + b_0 \lambda e^{-\lambda m} \right)$$

$$P(\beta; \sigma_\beta) = \frac{1}{\sqrt{2\pi}\sigma_\beta} e^{-\frac{\beta^2}{2\sigma_\beta^2}}$$

Systematic uncertainties

- The profile likelihood shape is broadened, with respect to to the usual likelihood function, due to the presence of nuisance parameter β (loss of information) that model systematic uncertainties
- Uncertainty on s increases
- Significance for discovery using s as test statistic decreases



This implementation is based on RooStats, a package, released as optional library with ROOT
http://root.cern.ch

Significance evaluation

- Assume $\mu = 0$, if $q_0 = -2 \ln \lambda(0)$ can be approximated by a χ^2 with one d.o.f., then the significance is approximately equal to:

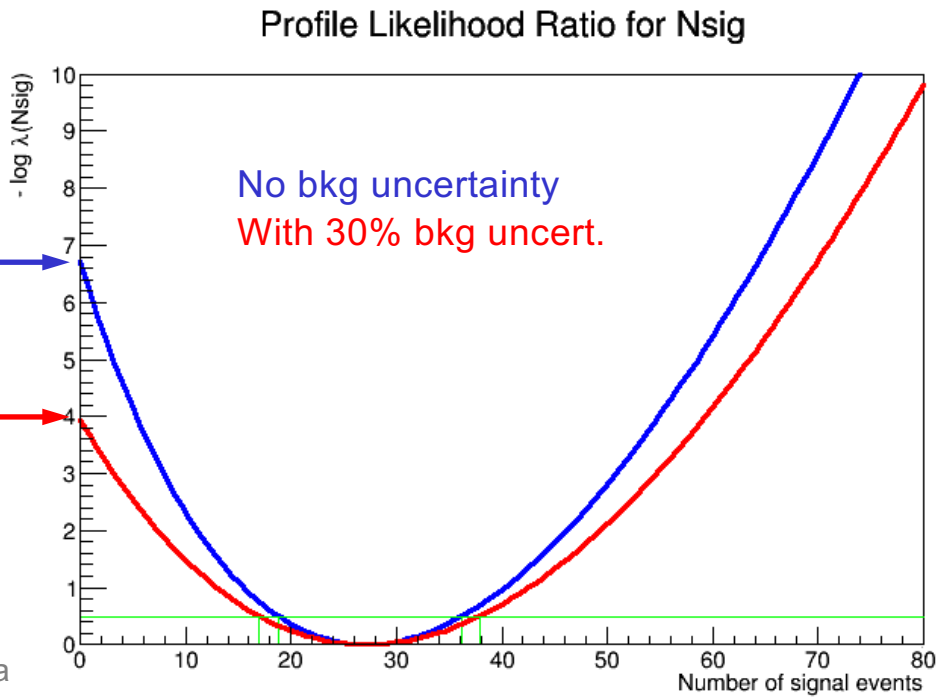
$$Z \cong \sqrt{q_0}$$

- The level of approximation can be verified with a computation done using pseudo experiments:
- Generate a large number of toy samples with zero background and determine the distribution of $q_0 = -2 \ln \lambda(0)$, then count the fraction of cases with values greater than the measured value (p-value), and convert it to Z:

$$Z = \Phi^{-1}(1 - p)$$

$$Z \cong \sqrt{2 \times 6.66} = 3.66$$

$$Z \cong \sqrt{2 \times 3.93} = 2.81$$



- Toy samples may be unpractical for very large Z

Variations on test statistic

G. Cowan et al., EPJ C71 (2011) 1554

- Test statistic for discovery:

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

  - In case of a negative estimate of μ, set the test statistic to zero: consider only positive μ as evidence against the background-only hypothesis. Approximately: $Z \cong \sqrt{q_0}$.

- Test statistic for upper limits:

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

  - If the estimate is larger than the assumed μ, an upward fluctuation occurred. Don't exclude μ in those cases, hence set the statistic to zero

- Higgs test statistic:

$$\tilde{q}_\mu = \begin{cases} -2\ln\dfrac{L(\vec{x}|\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(\vec{x}|0,\hat{\hat{\vec{\theta}}}(0))}\,, & \hat{\mu} < 0\,, \\ -2\ln\dfrac{L(\vec{x}|\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(\vec{x}|\hat{\mu},\hat{\hat{\vec{\theta}}})}\,, & 0 \leq \hat{\mu} \leq \mu\,, \\ 0\,, & \hat{\mu} > \mu\,. \end{cases}$$

← Protect for unphysical $\mu<0$

← As for upper limits statistic

LEP, Tevatron, LHC Higgs limits

	Test statistic	Profiled?	Test statistic sampling
LEP	$q_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\tilde{\theta})}{\mathcal{L}(data\|0,\tilde{\theta})}$	no	Bayesian-frequentist hybrid
Tevatron	$q_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|0,\hat{\theta}_0)}$	yes	Bayesian-frequentist hybrid
LHC	$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|\hat{\mu},\hat{\theta})}$	yes $(0 \leq \hat{\mu} \leq \mu)$	frequentist

Asymptotic approximations

- Asymptotic approximate formulae exist for most of adopted estimators
- If we want to test μ and we suppose data are distributed according to μ', we can write:

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$$

where $\hat{\mu}$ is distributed according to a Gaussian with average μ' and standard deviation σ (A. Wald, 1943)

- The covariance matrix can be asymptotically approximated by:

$$V_{ij}^{-1} = -\left\langle \frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j} \right\rangle$$

where μ' is assumed as signal strength value

- Case by case, the estimate of σ (from the inversion of V_{ij}^{-1}) can be determined

A. Wald, Trans. of AMS 54 n.3 (1943) 426-482

G. Cowan et al., EPJ C71 (2011) 1554

Asimov datasets

- Convenient to compute approximate values:
  *"We define the **Asimov data set** such that when one uses it to evaluate the estimators for all parameters, one obtains the true parameter values"*
- In practice: all observables are replaced with their expected value
- Yields expected values are possibly non integer

$$\lambda_A(\mu) = \frac{L_A(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_A(\hat{\mu}, \hat{\boldsymbol{\theta}})} = \frac{L_A(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_A(\mu', \boldsymbol{\theta})}$$

- Median significance for discovery or exclusion (and their $\pm 1\sigma$ bands) can be obtained using the Asimov dataset

$$\text{med}[Z_0|\mu'] = \sqrt{q_{0,A}}$$ ← For discovery using q_0

$$\text{med}[Z_\mu|0] = \sqrt{q_{\mu,A}}$$ ← For upper limit using q_μ

$$\text{med}[Z_\mu|0] = \sqrt{\tilde{q}_{\mu,A}}$$ ← Upper limits using $\tilde{q}_\mu$

In practice: all the interesting formulae are implemented in RooStats package, released as optional library in ROOT

G. Cowan et al., EPJ C71 (2011) 1554

The look-elsewhere effect

- Consider a search for a signal peak over a background distribution that is smoothly distributed over a wide range
- You could either:
  - Know which mass to look at, e.g.: search for a rare decay with a known particle, like $B_s \to \mu\mu$
  - Search for a peak at an unknown mass value, like for the Higgs boson
- In the former case it's easy to compute the peak significance:
  - Evaluate the test statistics for $\mu = 0$ (background only) at your observed data sample
  - Evaluate the p-value according to the expected distribution of your test statistic q under the background-only hypothesis, convert it to the equivalent area of a Gaussian tail to obtain the significance level:

$$p = \int_{q^{\mathrm{obs}}}^{\infty} f(q|\mu = 0)\mathrm{d}q \qquad Z = \Phi^{-1}(1-p)$$

The look-elsewhere effect

- In case you search for a peak at an unknown mass, the previous p-value has only a local meaning:
  - Probability to find a background fluctuation as large as your signal or more at a fixed mass value m:

$$p(m) = \int_{q^{\mathrm{obs}}(m)}^{\infty} f(q|\mu = 0)\mathrm{d}q$$

  - We need the probability to find a background fluctuation at least as large as your signal at **any** mass value (global)
  - local p-value would be an overestimate of the global p-value
- The chance that an over-fluctuation occurs on at least one mass value increases with the searched range
- Magnitude of the effect:
  - Roughly proportional to the ratio of resolution over the search range, also depending on the significance of the peak
  - Better resolution = less chance to have more events compatible with the same mass value
- Possible approach: let also m fluctuate in the test statistics fit:

$$\hat{q}_0 = -2\ln\frac{L(\mu = 0)}{L(\hat{\mu}; \hat{m})}$$

Note: for μ=0
L doesn't depend on m
Wilks' theorem doesn't apply

$$p^{\mathrm{glob}} = \int_{\hat{q}_0^{\mathrm{obs}}}^{\infty} f(\hat{q}_0|\mu = 0)\mathrm{d}\hat{q}_0$$

Estimate LEE

- The effect can be evaluated with brute-force Toy Monte Carlo:
  - Run N experiments with background-only
  - Find the maximum $\hat{q}$ of the test statistic q in the entire search range
  - Determine its distribution, hence compute the observed global p-value
  - Requires very large toy Monte Carlo samples (5σ: $p = 2.87 \times 10^{-7}$)

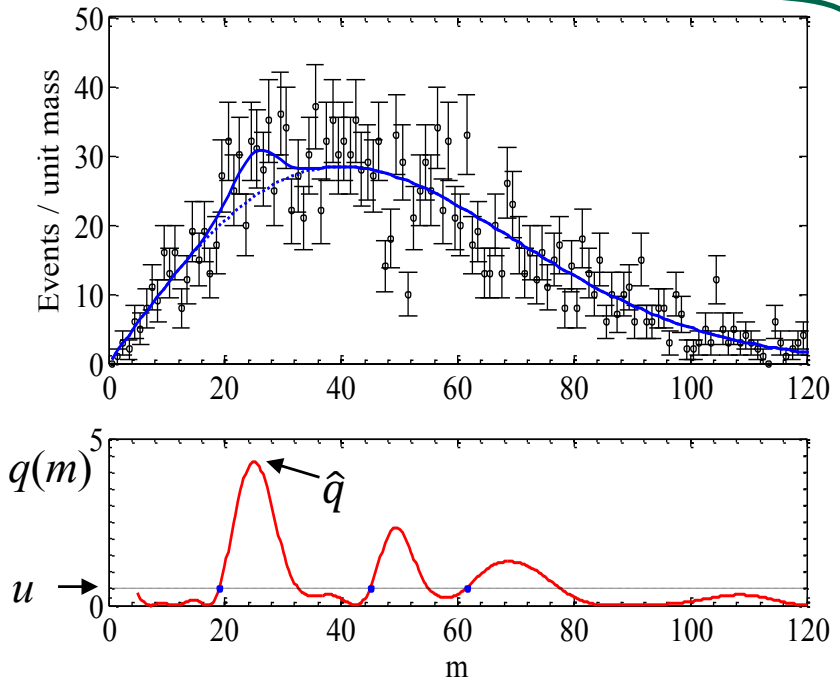
$$\hat{q} = \max_m q(m)$$

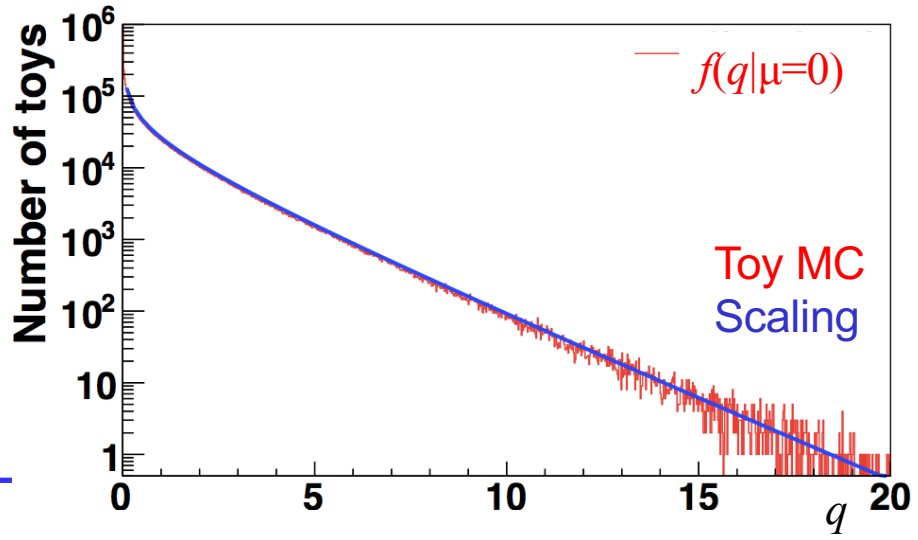
- Approximate evaluation based on local p-value, times correction factors ("trial factors", Gross and Vitells, EPJC 70:525-530,2010)

$$p^{\text{glob}} = P(\hat{q} > u) \leq \langle N_u \rangle + \frac{1}{2} P(\chi^2 > u)$$

$\langle N_u \rangle$ is the average number of up-crossings of the test statistic, can be evaluated at some lower reference level (toy MC) and scaled by:

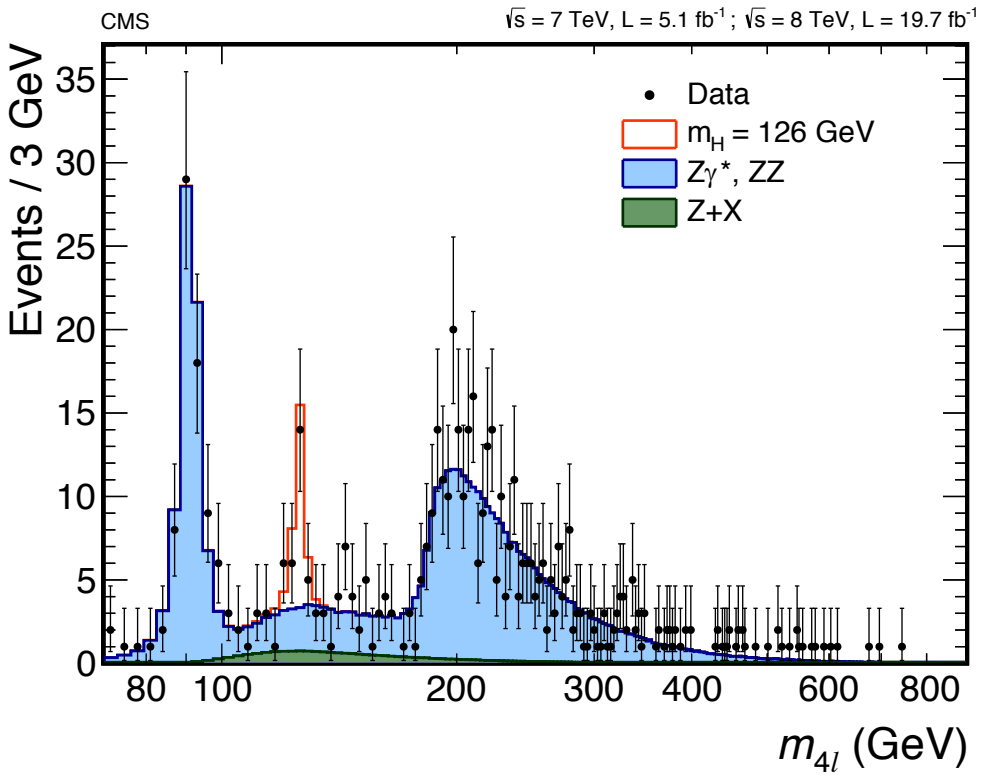
$$\langle N_u \rangle = \langle N_{u_0} \rangle \, e^{-\frac{u-u_0}{2}}$$

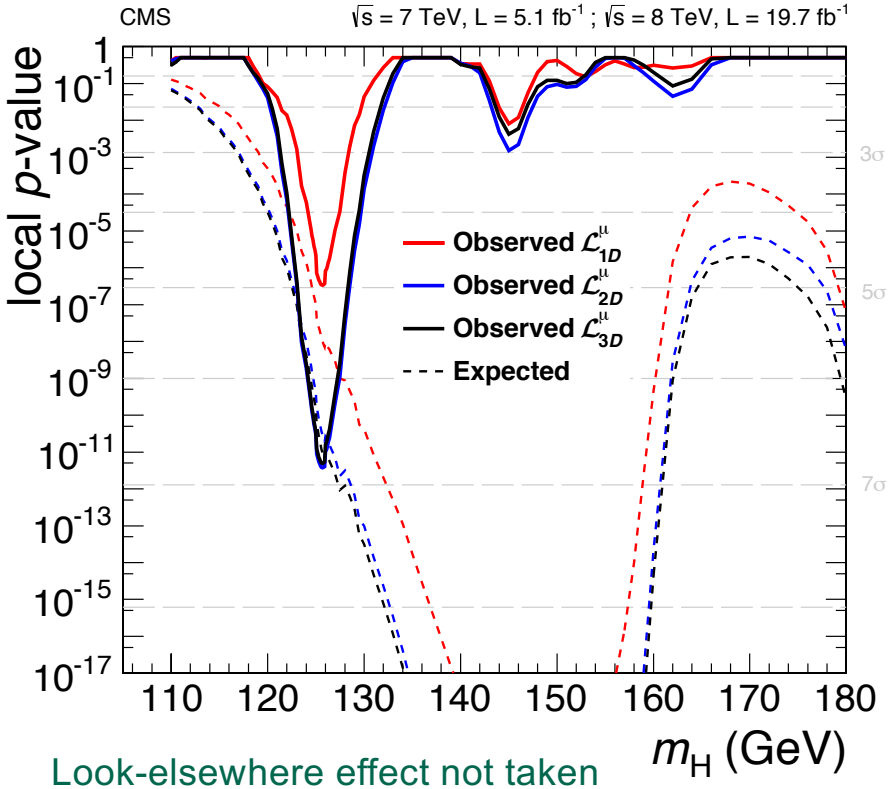




Putting all together

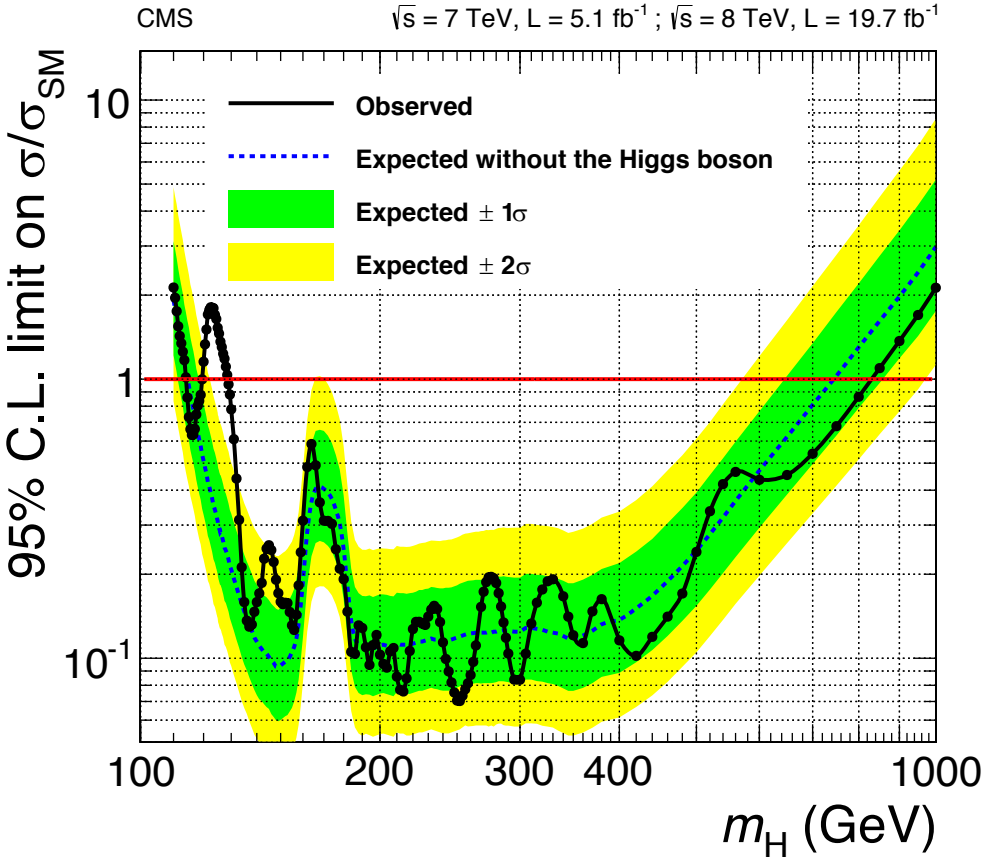
- Search for Higgs boson in H→4l at LHC
- 1D, 2D, 3D: different test statistics using 4l invariant mass plus other discriminating variables based on the event kinematics

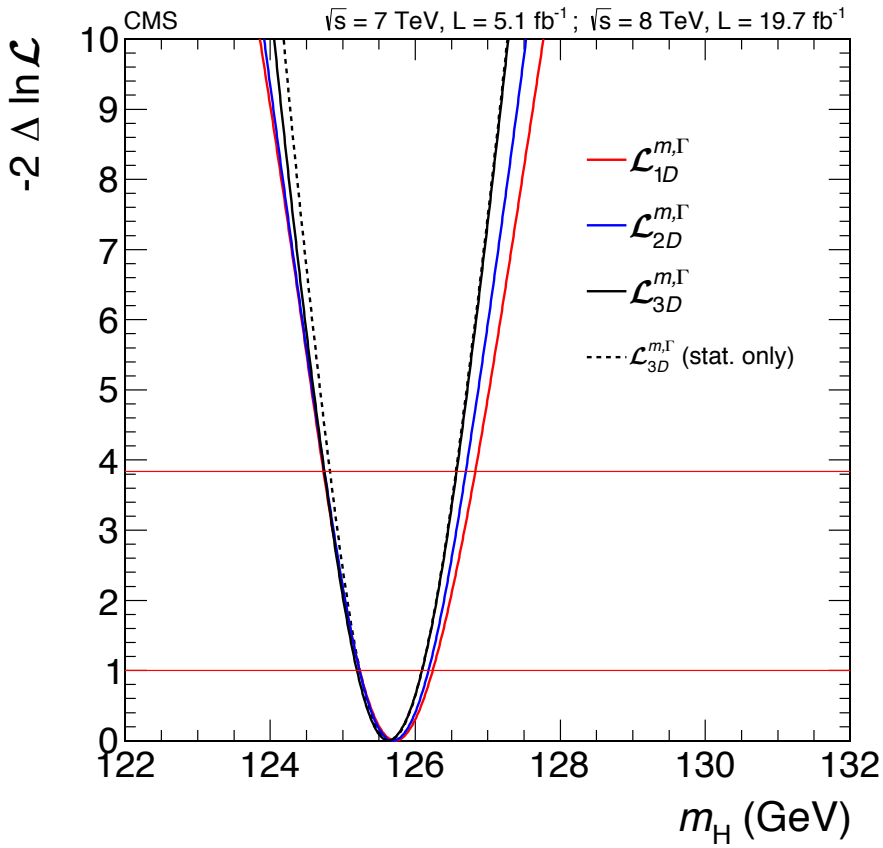




Look-elsewhere effect not taken into account here

Higgs exclusion





"The modified frequentist construction CLs is adopted as the primary method for reporting limits. As a complementary method to the frequentist construction, a Bayesian approach yields consistent results."

Agreed statistical procedure described in:
ATLAS and CMS Collaborations,
LHC Higgs Combination Group
ATL-PHYS-PUB 2011-11/CMS NOTE
2011/005, 2011.

The End

- What will be next discovery?