Statistics for HEP

Lecture 2: Discovery and Limits

http://indico.cern.ch/conferenceDisplay.py?confId=173727

Academic Training Lectures

CERN, 2–5 April, 2012

Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Outline

Lecture 1: Introduction and basic formalism
- Probability, statistical tests, parameter estimation.

→ Lecture 2: Discovery and Limits
- Quantifying discovery significance and sensitivity
- Frequentist intervals/limits

Lecture 3: More on discovery and limits
- Bayesian intervals/limits
- The Look-Elsewhere Effect
- Dealing with nuisance parameters

Lecture 4: Unfolding (deconvolution)
- Correcting distributions for effects of smearing

Reminder about statistical tests

Consider test of a parameter μ, e.g., proportional to signal rate.

Result of measurement is a set of numbers $\mathbf{x}$.

To define test of μ, specify *critical region* w_μ, such that probability to find $\mathbf{x} \in w_\mu$ is not greater than α (the *size* or *significance level*):

$$P(\mathbf{x} \in w_\mu | \mu) \leq \alpha$$

(Must use inequality since $\mathbf{x}$ may be discrete, so there may not exist a subset of the data space with probability of exactly α.)

Equivalently define a *p*-value p_μ such that the critical region corresponds to $p_\mu \leq \alpha$.

Often use, e.g., $\alpha = 0.05$.

If observe $\mathbf{x} \in w_\mu$, reject μ.

Confidence interval from inversion of a test

Carry out a test of size α for all values of μ.

The values that are not rejected constitute a *confidence interval* for μ at confidence level CL = $1 - \alpha$.

> The probability that the true value of μ will be rejected is not greater than α, so by construction the confidence interval will contain the true value of μ with probability $\geq 1 - \alpha$.

The interval depends on the choice of the test (critical region).

If the test is formulated in terms of a p-value, p_μ, then the confidence interval represents those values of μ for which $p_\mu > \alpha$.

To find the end points of the interval, set $p_\mu = \alpha$ and solve for μ.

Choice of test for discovery

If μ represents the signal rate, then discovering the signal process requires rejecting $H_0 : \mu = 0$.

Often our evidence for the signal process comes in the form of an excess of events above the level predicted from background alone, i.e., $\mu > 0$ for physical signal models.

So the relevant alternative hypothesis is $H_0 : \mu > 0$.

In other cases the relevant alternative may also include $\mu < 0$ (e.g., neutrino oscillations).

The critical region giving the highest power for the test of $\mu = 0$ relative to the alternative of $\mu > 0$ thus contains high values of the estimated signal rate.

Choice of test for limits

Suppose the existence of the signal process ($\mu > 0$) is not yet established.

The interesting alternative in this context is $\mu = 0$.

That is, we want to ask what values of μ can be excluded on the grounds that the implied rate is too high relative to what is observed in the data.

The critical region giving the highest power for the test of μ relative to the alternative of $\mu = 0$ thus contains low values of the estimated rate, $\hat{\mu}$.

Test based on one-sided alternative → upper limit.

More on choice of test for limits

In other cases we want to exclude μ on the grounds that some other measure of incompatibility between it and the data exceeds some threshold.

For example, the process may be known to exist, and thus $\mu = 0$ is no longer an interesting alternative.

If the measure of incompatibility is taken to be the likelihood ratio with respect to a two-sided alternative, then the critical region can contain data values corresponding to both high and low signal rate.

→ unified intervals, G. Feldman, R. Cousins,
Phys. Rev. D 57, 3873–3889 (1998)

The Big Debate is whether to focus on small (or zero) values of the parameter as the relevant alternative when the existence of a signal has not yet been established. Professional statisticians have voiced support on both sides of the debate.

A simple example

For each event we measure two variables, $\boldsymbol{x} = (x_1, x_2)$.

Suppose that for background events (hypothesis H_0),

$$f(\mathbf{x}|H_0) = \frac{1}{\xi_1} e^{-x_1/\xi_1} \, \frac{1}{\xi_2} e^{-x_2/\xi_2}$$

and for a certain signal model (hypothesis H_1) they follow

$$f(\mathbf{x}|H_1) = C \, \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x_1-\mu_1)^2/2\sigma_1^2} \, \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x_2-\mu_2)^2/2\sigma_2^2}$$

where $x_1, x_2 \geq 0$ and C is a normalization constant.

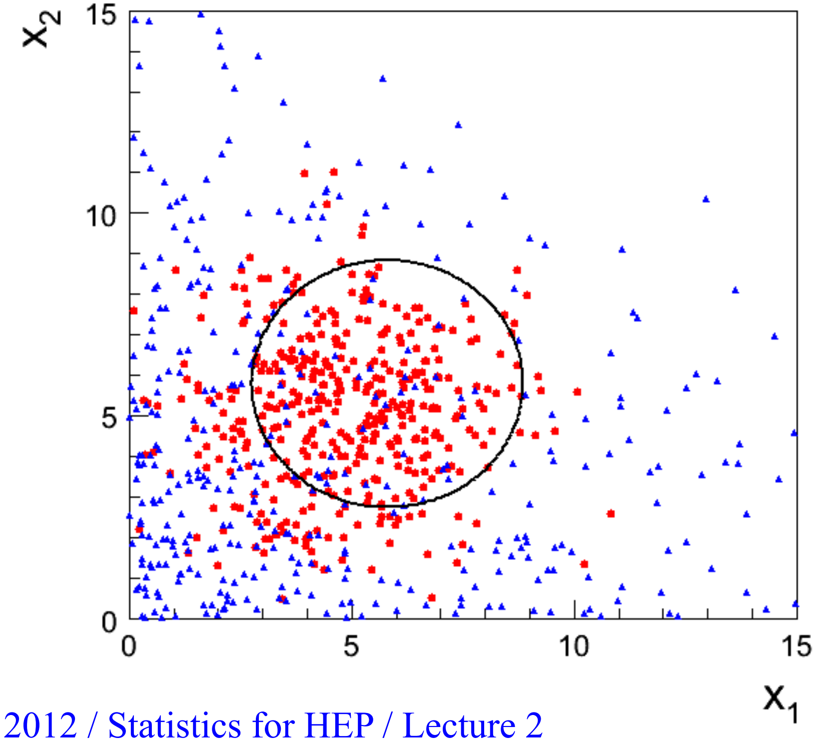
Likelihood ratio as test statistic

In a real-world problem we usually wouldn't have the pdfs $f(\boldsymbol{x}|H_0)$ and $f(\boldsymbol{x}|H_1)$, so we wouldn't be able to evaluate the likelihood ratio

$$t(\mathbf{x}) = \frac{f(\mathbf{x}|H_1)}{f(\mathbf{x}|H_0)}$$

for a given observed $\boldsymbol{x}$, hence the need for multivariate methods to approximate this with some other function.

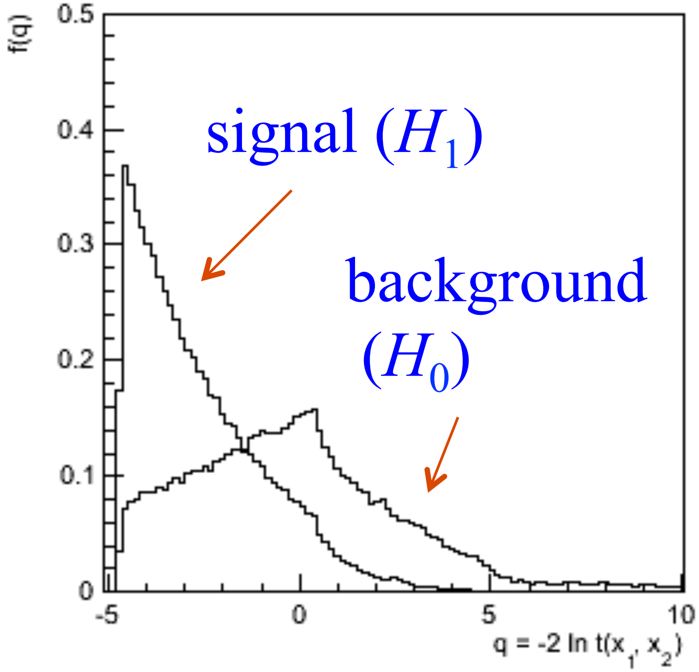
But in this example we can find contours of constant likelihood ratio such as:



Event selection using the LR

Using Monte Carlo, we can find the distribution of the likelihood ratio or equivalently of

$$q = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - \frac{2x_1}{\xi_1} - \frac{2x_2}{\xi_2} = -2\ln t(\mathbf{x}) + C$$



signal (H_1)

background (H_0)

From the Neyman-Pearson lemma we know that by cutting on this variable we would select a signal sample with the highest signal efficiency (test power) for a given background efficiency.

Search for the signal process

But what if the signal process is not known to exist and we want to search for it. The relevant hypotheses are therefore

H_0: all events are of the background type
H_1: the events are a mixture of signal and background

Rejecting H_0 with $Z > 5$ constitutes "discovering" new physics.

Suppose that for a given integrated luminosity, the expected number of signal events is s, and for background b.

The observed number of events n will follow a Poisson distribution:

$$P(n|b) = \frac{b^n}{n!} e^{-b} \qquad P(n|s+b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

Likelihoods for full experiment

We observe n events, and thus measure n instances of $\boldsymbol{x} = (x_1, x_2)$.

The likelihood function for the entire experiment assuming the background-only hypothesis (H_0) is

$$L_b = \frac{b^n}{n!} e^{-b} \prod_{i=1}^{n} f(\mathbf{x}_i | \mathrm{b})$$

and for the "signal plus background" hypothesis (H_1) it is

$$L_{s+b} = \frac{(s+b)^n}{n!} e^{-(s+b)} \prod_{i=1}^{n} \left(\pi_{\mathrm{s}} f(\mathbf{x}_i | \mathrm{s}) + \pi_{\mathrm{b}} f(\mathbf{x}_i | \mathrm{b}) \right)$$

where π_s and π_b are the (prior) probabilities for an event to be signal or background, respectively.

Likelihood ratio for full experiment

We can define a test statistic Q monotonic in the likelihood ratio as

$$Q = -2 \ln \frac{L_{s+b}}{L_b} = -s + \sum_{i=1}^{n} \ln \left(1 + \frac{s}{b} \frac{f(\mathbf{x}_i|\mathrm{s})}{f(\mathbf{x}_i|\mathrm{b})} \right)$$

To compute *p*-values for the b and s+b hypotheses given an observed value of Q we need the distributions $f(Q|\mathrm{b})$ and $f(Q|\mathrm{s+b})$.

Note that the term $-s$ in front is a constant and can be dropped.

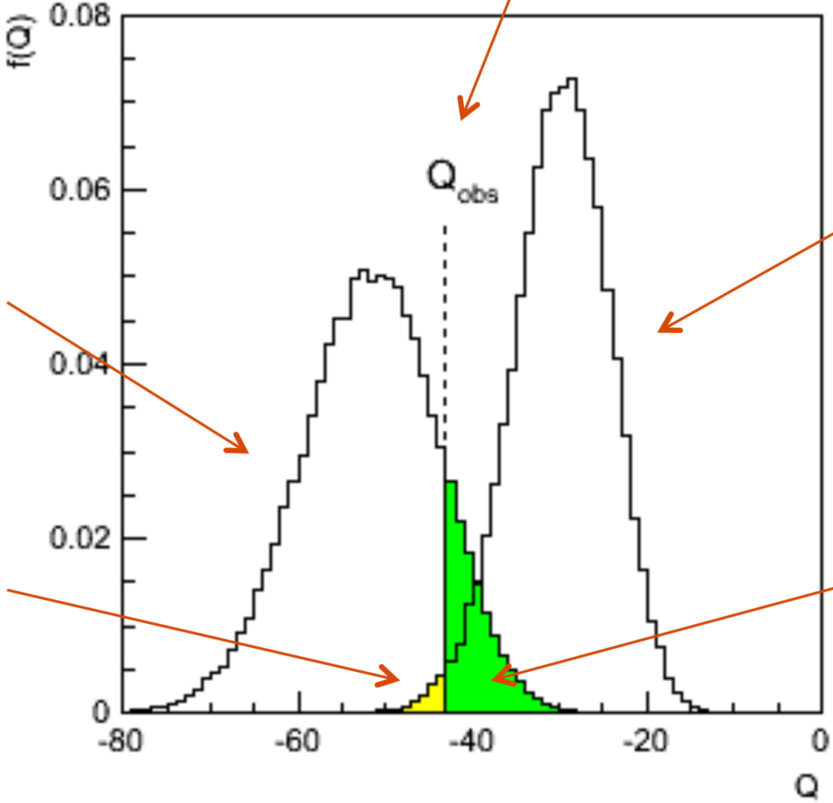
The rest is a sum of contributions for each event, and each term in the sum has the same distribution.

Can exploit this to relate distribution of Q to that of single event terms using (Fast) Fourier Transforms (Hu and Nielsen, physics/9906010).

Distribution of Q

Take e.g. $b = 100$, $s = 20$.

Suppose in real experiment Q is observed here.



$f\,(Q|\mathrm{s+b})$

$f\,(Q|\mathrm{b})$

p-value of b only

p-value of s+b

If $p_{s+b} < \alpha$, reject signal model s at confidence level $1 - \alpha$.

If $p_b < 2.9 \times 10^{-7}$, reject background-only model (signif. $Z = 5$).

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

Large-sample approximations for prototype analysis using profile likelihood ratio

Search for signal in a region of phase space; result is histogram of some variable *x* giving numbers:

$$\mathbf{n} = (n_1, \ldots, n_N)$$

Assume the n_i are Poisson distributed with expectation values

$$E[n_i] = \mu s_i + b_i$$

strength parameter

where

$$s_i = s_{\rm tot} \int_{\mathrm{bin}\ i} f_s(x; \boldsymbol{\theta}_s)\, dx \,, \qquad b_i = b_{\rm tot} \int_{\mathrm{bin}\ i} f_b(x; \boldsymbol{\theta}_b)\, dx \,.$$

signal

background

Prototype analysis (II)

Often also have a subsidiary measurement that constrains some of the background and/or shape parameters:

$$\mathbf{m} = (m_1, \ldots, m_M)$$

Assume the m_i are Poisson distributed with expectation values

$$E[m_i] = u_i(\boldsymbol{\theta})$$

nuisance parameters ($\boldsymbol{\theta}_{\mathrm{s}}$, $\boldsymbol{\theta}_{\mathrm{b}}$,$b_{\mathrm{tot}}$)

Likelihood function is

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \quad \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

The profile likelihood ratio

Base significance test on the profile likelihood ratio:

maximizes *L* for
Specified *μ*

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

maximize *L*

The likelihood ratio of point hypotheses gives optimum test (Neyman-Pearson lemma).

The profile LR in the present analysis with variable μ and nuisance parameters $\boldsymbol{\theta}$ is expected to be near optimal.

Test statistic for discovery

Try to reject background-only ($\mu = 0$) hypothesis using

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

i.e. here only regard upward fluctuation of data as evidence against the background-only hypothesis.

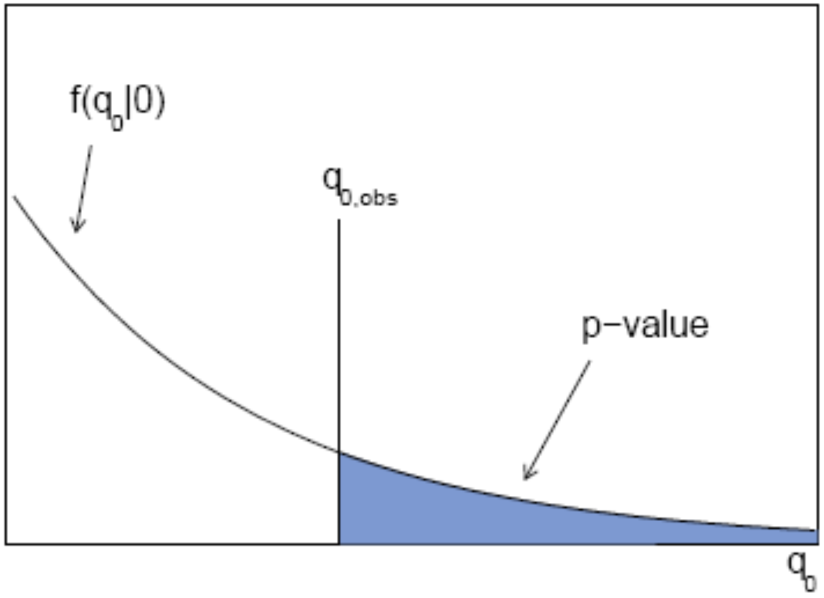
Note that even though here physically $\mu \geq 0$, we allow $\hat{\mu}$ to be negative. In large sample limit its distribution becomes Gaussian, and this will allow us to write down simple expressions for distributions of our test statistics.

p-value for discovery

Large q_0 means increasing incompatibility between the data and hypothesis, therefore p-value for an observed $q_{0,\mathrm{obs}}$ is

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0$$

will get formula for this later

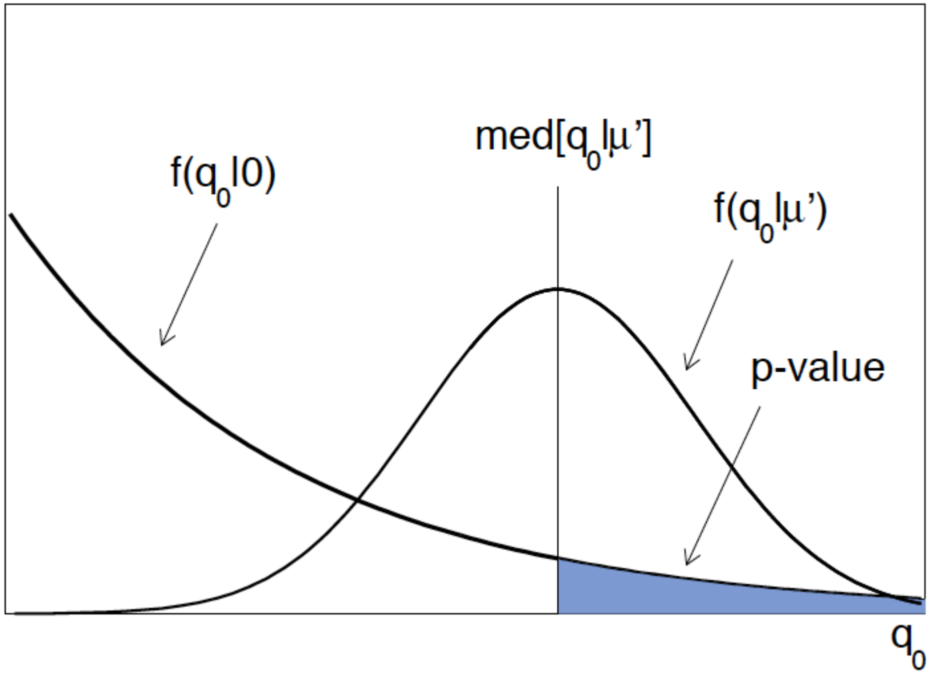


From p-value get equivalent significance,

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

Expected (or median) significance / sensitivity

When planning the experiment, we want to quantify how sensitive we are to a potential discovery, e.g., by given median significance assuming some nonzero strength parameter μ'.



So for *p*-value, need $f(q_0|0)$, for sensitivity, will need $f(q_0|\mu')$,

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

Distribution of q_0 in large-sample limit

Assuming approximations valid in the large sample (asymptotic) limit, we can write down the full distribution of q_0 as

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

The special case $\mu' = 0$ is a "half chi-square" distribution:

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}$$

In large sample limit, $f(q_0|0)$ independent of nuisance parameters; $f(q_0|\mu')$ depends on nuisance parameters through σ.

Cumulative distribution of q_0, significance

From the pdf, the cumulative distribution of q_0 is found to be

$$F(q_0|\mu') = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)$$

The special case $\mu' = 0$ is

$$F(q_0|0) = \Phi\left(\sqrt{q_0}\right)$$

The *p*-value of the $\mu = 0$ hypothesis is

$$p_0 = 1 - F(q_0|0)$$

Therefore the discovery significance Z is simply

$$\boxed{Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}}$$

Test statistic for upper limits

For purposes of setting an upper limit on μ one may use

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases} \qquad \text{where} \qquad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

Note for purposes of setting an upper limit, one does not regard an upwards fluctuation of the data as representing incompatibility with the hypothesized μ.

From observed q_μ find p-value: $$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu)\, dq_\mu$$

95% CL upper limit on μ is highest value for which p-value is not less than 0.05.

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

Distribution of q_μ in large-sample limit

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu' - \mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)^2\right]$$

$$f(q_\mu|\mu) = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2}$$

$$F(q_\mu|\mu') = \Phi\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)$$

$$p_\mu = 1 - F(q_\mu|\mu) = 1 - \Phi\left(\sqrt{q_\mu}\right)$$

Independent of nuisance parameters.

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

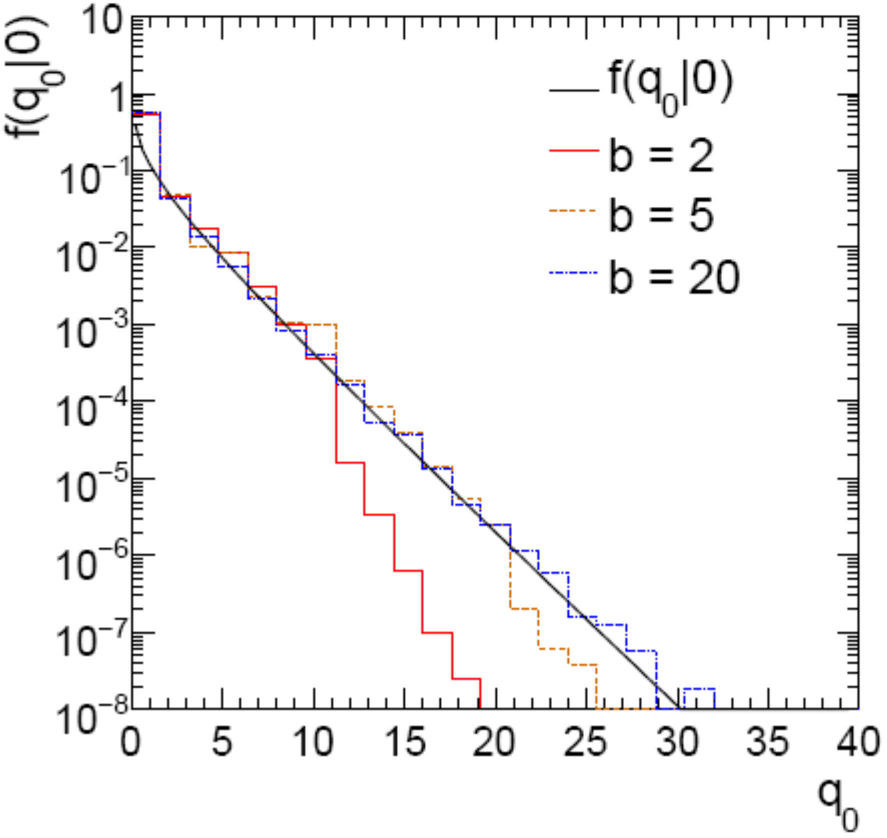
Monte Carlo test of asymptotic formula

$$n \sim \text{Poisson}(\mu s + b)$$

$$m \sim \text{Poisson}(\tau b)$$

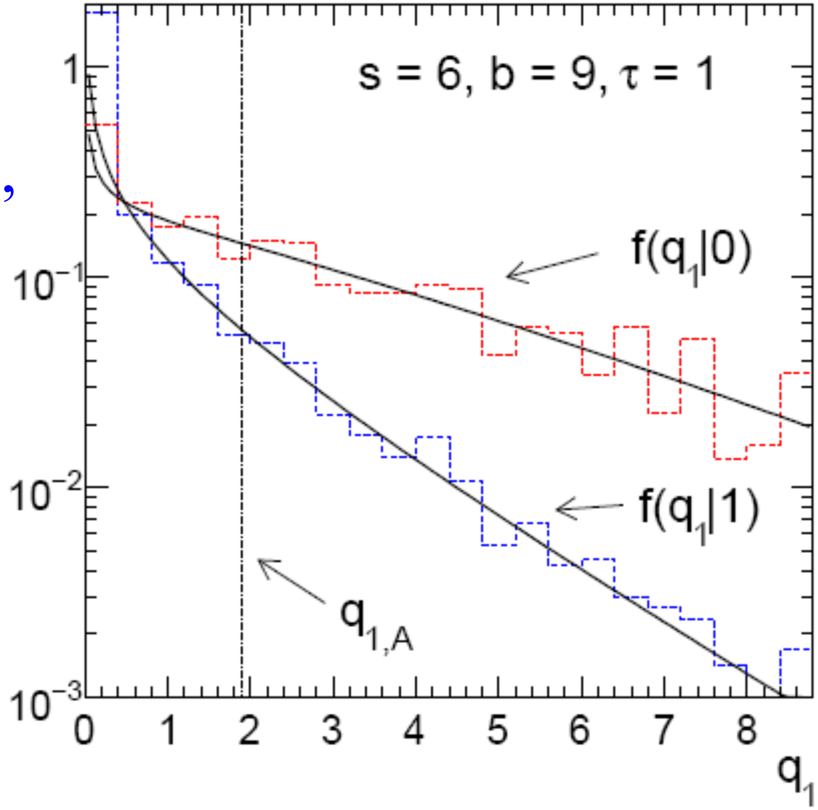
Here take $\tau = 1$.

Asymptotic formula is good approximation to 5σ level ($q_0 = 25$) already for $b \sim 20$.



Monte Carlo test of asymptotic formulae

Consider again $n \sim$ Poisson $(\mu s + b)$, $m \sim$ Poisson(τb)
Use q_μ to find p-value of hypothesized μ values.



E.g. $f(q_1|1)$ for p-value of $\mu = 1$.

Typically interested in 95% CL, i.e., p-value threshold = 0.05, i.e., $q_1 = 2.69$ or $Z_1 = \sqrt{q_1} = 1.64$.

Median$[q_1|0]$ gives "exclusion sensitivity".

Here asymptotic formulae good for $s = 6$, $b = 9$.

Unified (Feldman-Cousins) intervals

We can use directly

$$t_\mu = -2\ln\lambda(\mu) \qquad \text{where} \qquad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}$$

as a test statistic for a hypothesized μ.

Large discrepancy between data and hypothesis can correspond either to the estimate for μ being observed high or low relative to μ.

This is essentially the statistic used for Feldman-Cousins intervals (here also treats nuisance parameters).

G. Feldman and R.D. Cousins, Phys. Rev. D 57 (1998) 3873.

Distribution of t_μ

Using Wald approximation, $f(t_\mu|\mu')$ is noncentral chi-square for one degree of freedom:

$$f(t_\mu|\mu') = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\frac{\mu-\mu'}{\sigma}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right)\right]$$

Special case of $\mu = \mu'$ is chi-square for one d.o.f. (Wilks).

The p-value for an observed value of t_μ is

$$p_\mu = 1 - F(t_\mu|\mu) = 2\left(1-\Phi\left(\sqrt{t_\mu}\right)\right)$$

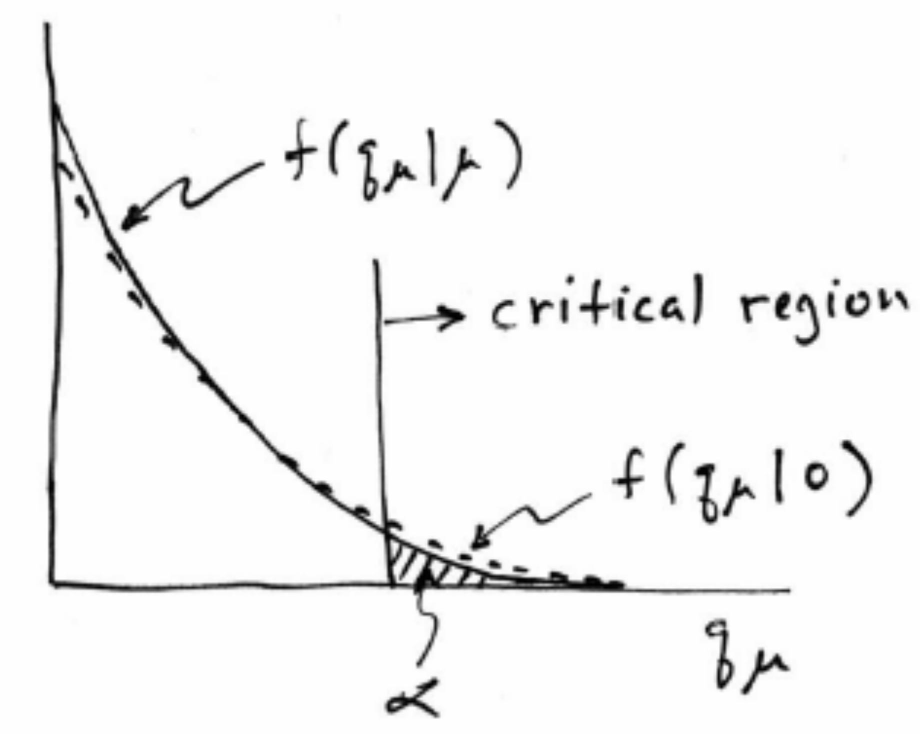
and the corresponding significance is

$$Z_\mu = \Phi^{-1}(1-p_\mu) = \Phi^{-1}\left(2\Phi\left(\sqrt{t_\mu}\right)-1\right)$$

Low sensitivity to μ

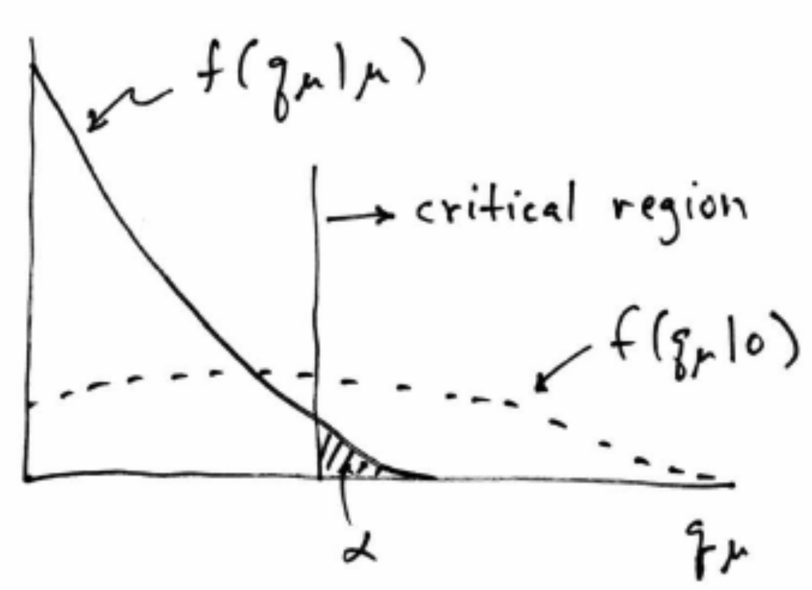
It can be that the effect of a given hypothesized μ is very small relative to the background-only ($\mu = 0$) prediction.

This means that the distributions $f(q_\mu|\mu)$ and $f(q_\mu|0)$ will be almost the same:



Having sufficient sensitivity

In contrast, having sensitivity to μ means that the distributions $f(q_\mu|\mu)$ and $f(q_\mu|0)$ are more separated:

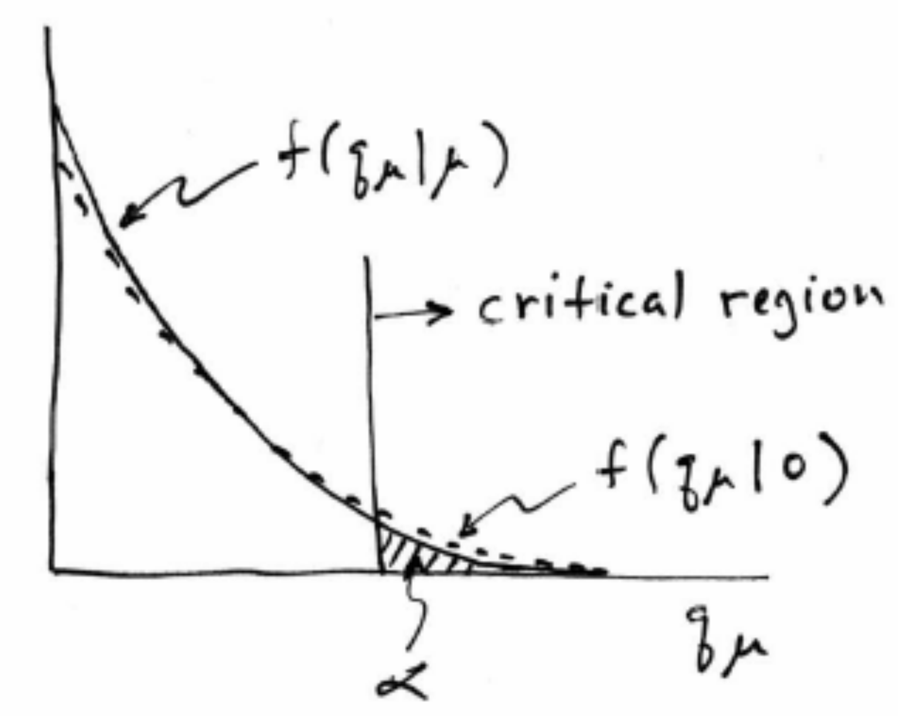


That is, the power (probability to reject μ if $\mu = 0$) is substantially higher than α. Use this power as a measure of the sensitivity.

Spurious exclusion

Consider again the case of low sensitivity. By construction the probability to reject μ if μ is true is α (e.g., 5%).

And the probability to reject μ if $\mu = 0$ (the power) is only slightly greater than α.

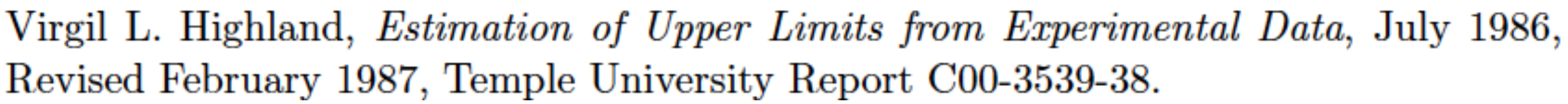


This means that with probability of around $\alpha = 5\%$ (slightly higher), one excludes hypotheses to which one has essentially no sensitivity (e.g., $m_{\mathrm{H}} = 1000$ TeV).

“Spurious exclusion”

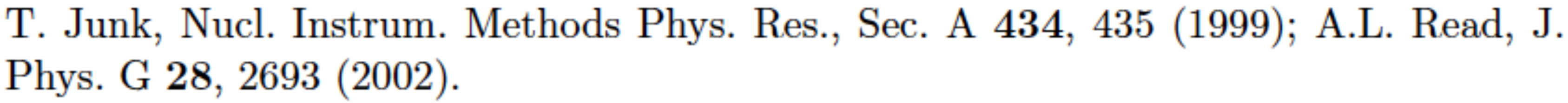
Ways of addressing spurious exclusion

The problem of excluding parameter values to which one has no sensitivity known for a long time; see e.g.,



Virgil L. Highland, *Estimation of Upper Limits from Experimental Data*, July 1986, Revised February 1987, Temple University Report C00-3539-38.

In the 1990s this was re-examined for the LEP Higgs search by Alex Read and others



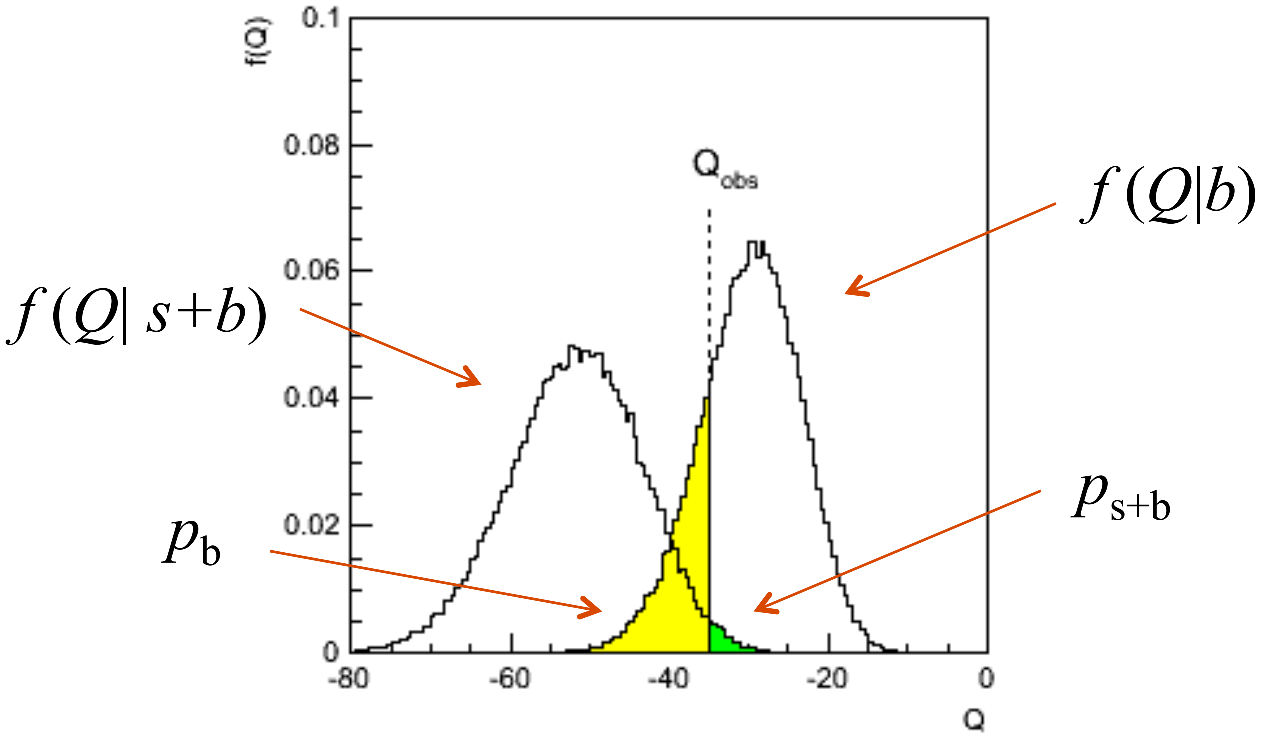
T. Junk, Nucl. Instrum. Methods Phys. Res., Sec. A **434**, 435 (1999); A.L. Read, J. Phys. G **28**, 2693 (2002).

and led to the "CL$_s$" procedure for upper limits.

Unified intervals also effectively reduce spurious exclusion by the particular choice of critical region.

The CL$_s$ procedure

In the usual formulation of CL$_s$, one tests both the $\mu = 0$ (b) and $\mu > 0$ (μs+b) hypotheses with the same statistic $Q = -2\ln L_{s+b}/L_b$:



$f\,(Q|\,s$+$b)$

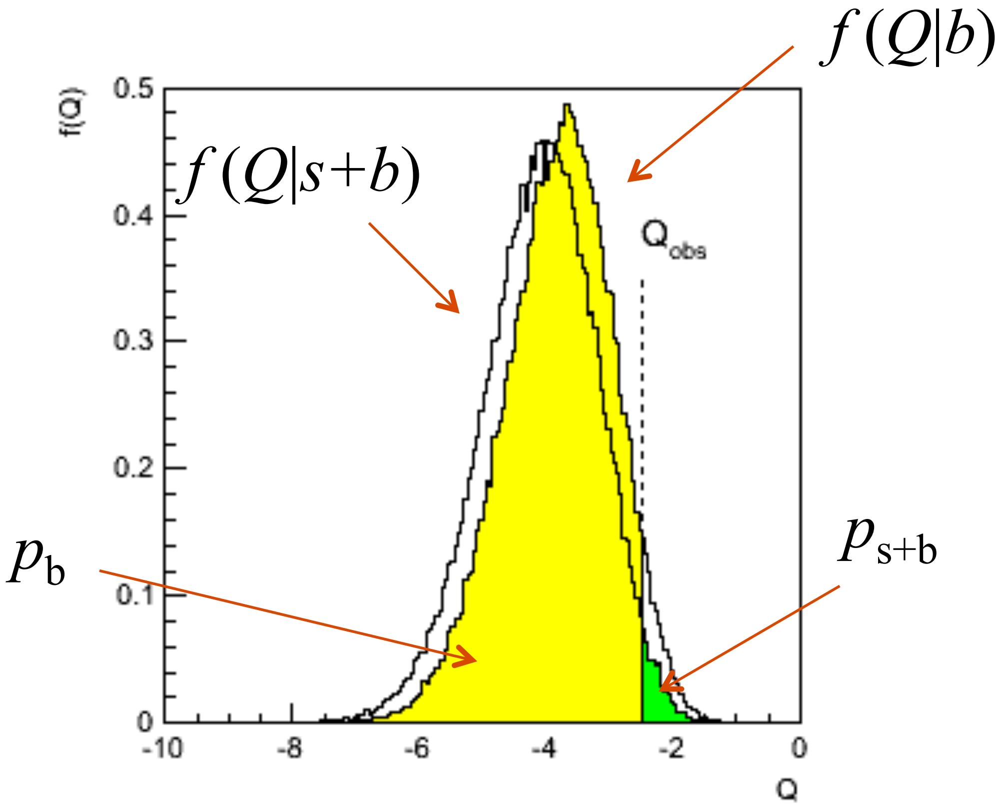
$f\,(Q|b)$

$p_{\rm b}$

$p_{\rm s+b}$

The CL_s procedure (2)

As before, “low sensitivity” means the distributions of Q under b and $s+b$ are very close:



The CL$_s$ procedure (3)

The CL$_\mathrm{s}$ solution (A. Read et al.) is to base the test not on the usual *p*-value (CL$_\mathrm{s+b}$), but rather to divide this by CL$_\mathrm{b}$ (~ one minus the *p*-value of the *b*-only hypothesis), i.e.,

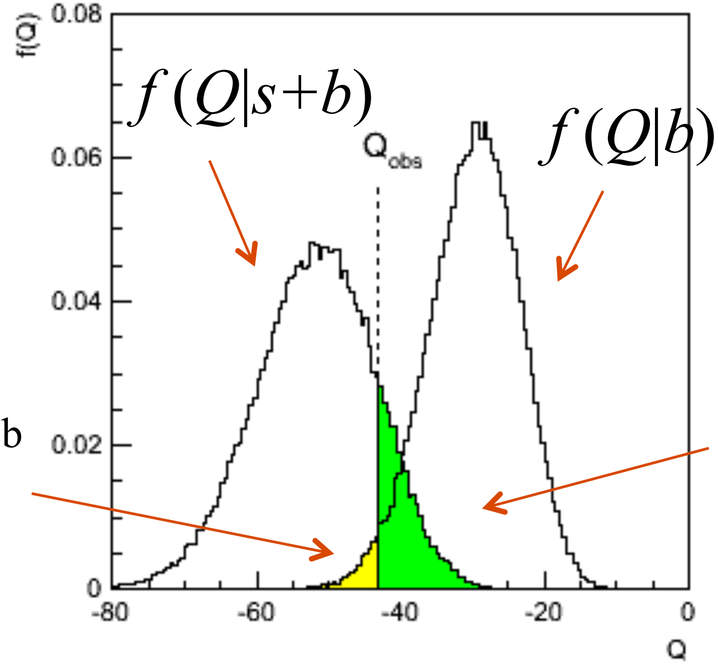
Define:

$$\mathrm{CL_s} = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}}$$

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

Reject s+b hypothesis if:

$$\mathrm{CL_s} \le \alpha$$



$1-\mathrm{CL_b}$
$= p_\mathrm{b}$

$\mathrm{CL_{s+b}}$
$= p_\mathrm{s+b}$

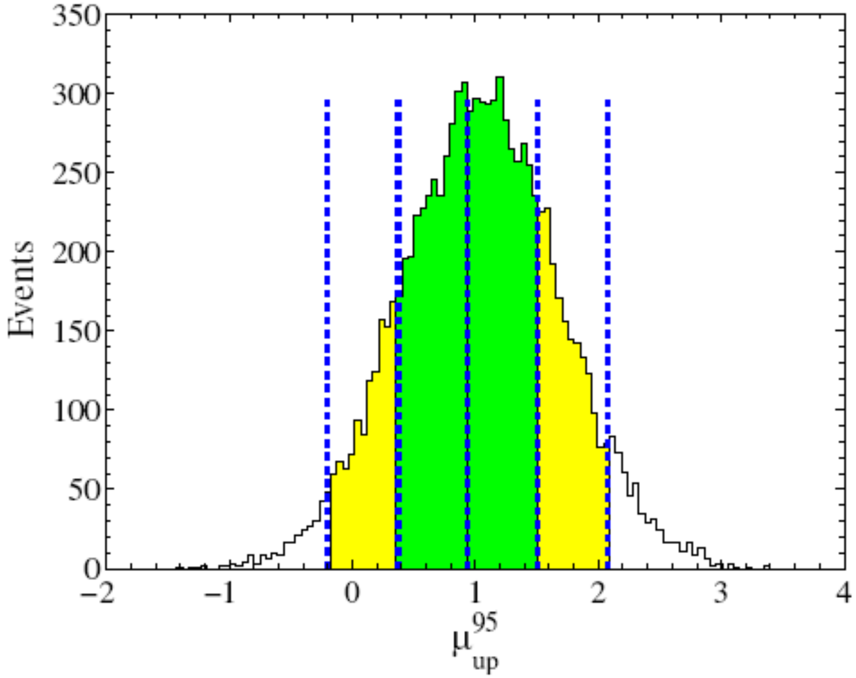
Reduces "effective" *p*-value when the two distributions become close (prevents exclusion if sensitivity is low).

Setting upper limits on $\mu = \sigma/\sigma_{SM}$

Carry out the CL"s" procedure for the parameter $\mu = \sigma/\sigma_{SM}$, resulting in an upper limit μ_{up}.

In, e.g., a Higgs search, this is done for each value of m_H.

At a given value of m_H, we have an observed value of μ_{up}, and we can also find the distribution $f(\mu_{up}|0)$:



±1σ (green) and ±2σ (yellow) bands from toy MC;

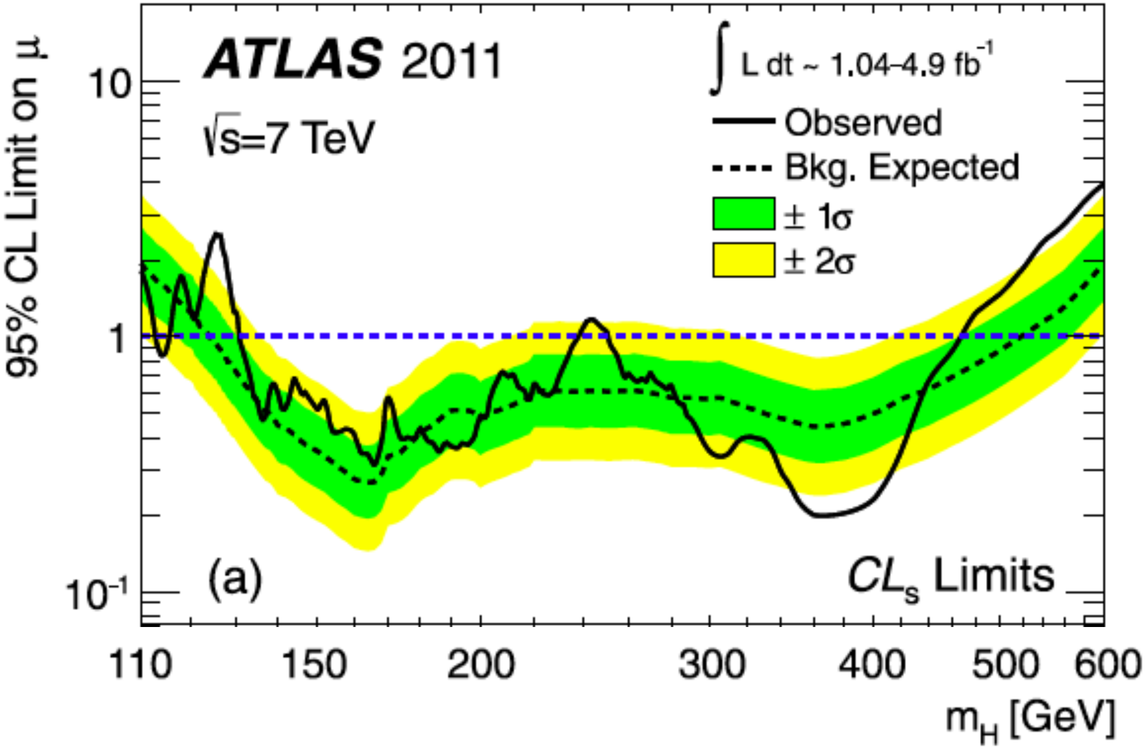
Vertical lines from asymptotic formulae.

How to read the green and yellow limit plots

For every value of m_{H}, find the CLs upper limit on μ.

Also for each m_{H}, determine the distribution of upper limits μ_{up} one would obtain under the hypothesis of $\mu = 0$.

The dashed curve is the median μ_{up}, and the green (yellow) bands give the $\pm 1\sigma$ (2σ) regions of this distribution.

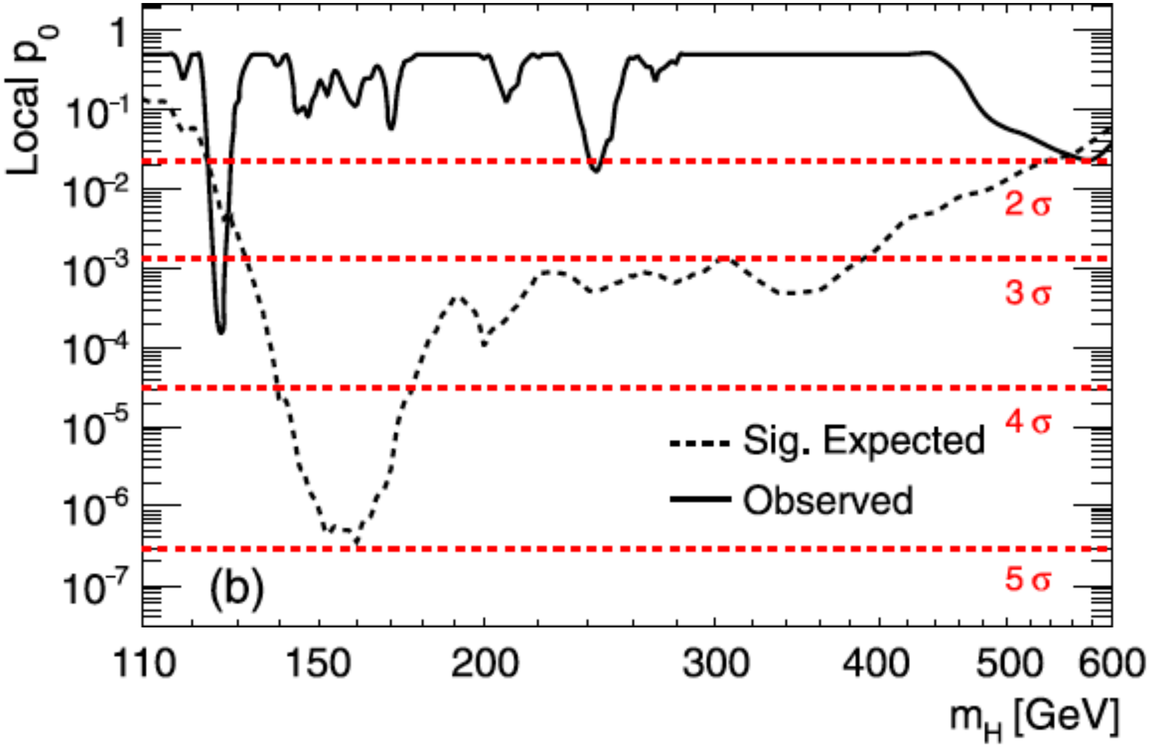


ATLAS, Phys. Lett. B 710 (2012) 49-66

How to read the p_0 plot

The "local" p_0 means the p-value of the background-only hypothesis obtained from the test of $\mu = 0$ at each individual m_H, without any correct for the Look-Elsewhere Effect.

The "Sig. Expected" (dashed) curve gives the median p_0 under assumption of the SM Higgs ($\mu = 1$) at each m_H.



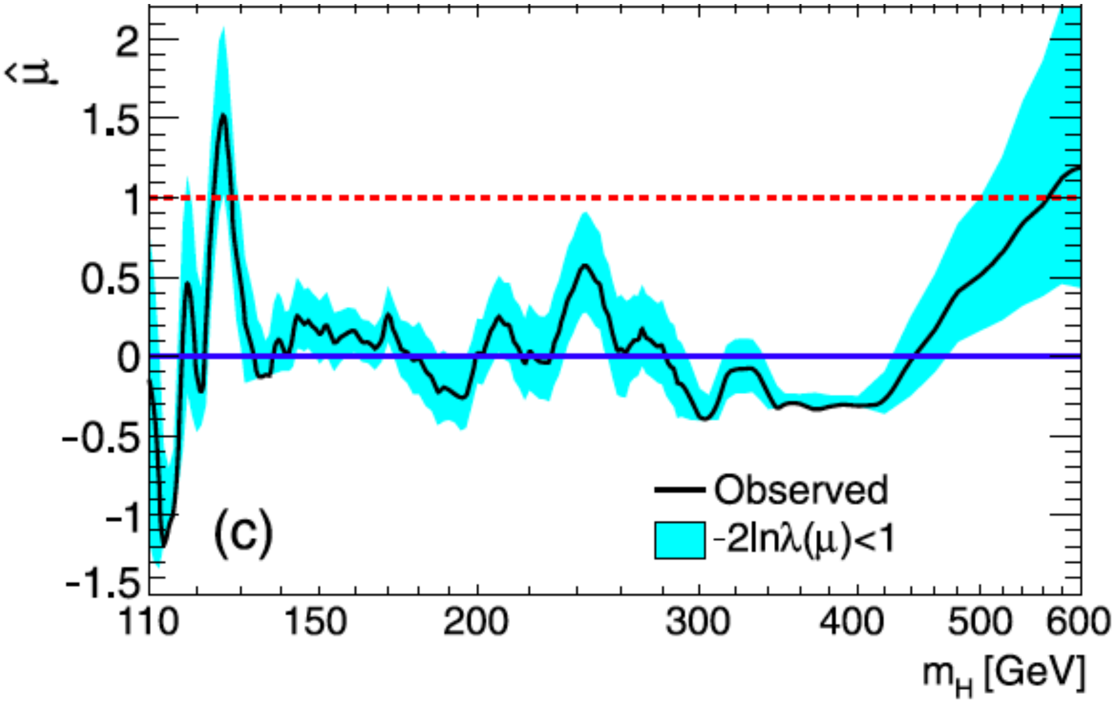
ATLAS, Phys. Lett.
B 710 (2012) 49-66

How to read the “blue band”

On the plot of $\hat{\mu}$ versus m_H, the blue band is defined by

$$-2\ln\lambda(\mu) = -2\ln(L(\mu)/L(\hat{\mu})) < 1 \text{ i.e., } \ln L(\mu) > \ln L(\hat{\mu}) - \frac{1}{2}$$

i.e., it approximates the 1-sigma error band (68.3% CL conf. int.)



ATLAS, Phys. Lett.
B 710 (2012) 49-66

Summary of Lecture 2

Confidence intervals obtained from inversion of a test of all parameter values.

Freedom to choose e.g. one- or two-sided test, often based on a likelihood ratio statistic.

Distributions of likelihood-ratio statistics can be written down in simple form for large-sample (asymptotic) limit.

Usual procedure for upper limit based on one-sided test can reject parameter values to which one has no sensitivity.

Various solutions; so far we have seen CLs.

Extra slides

Alternative test statistic for upper limits

Assume physical signal model has $\mu > 0$, therefore if estimator for μ comes out negative, the closest physical model has $\mu = 0$.

Therefore could also measure level of discrepancy between data and hypothesized μ with

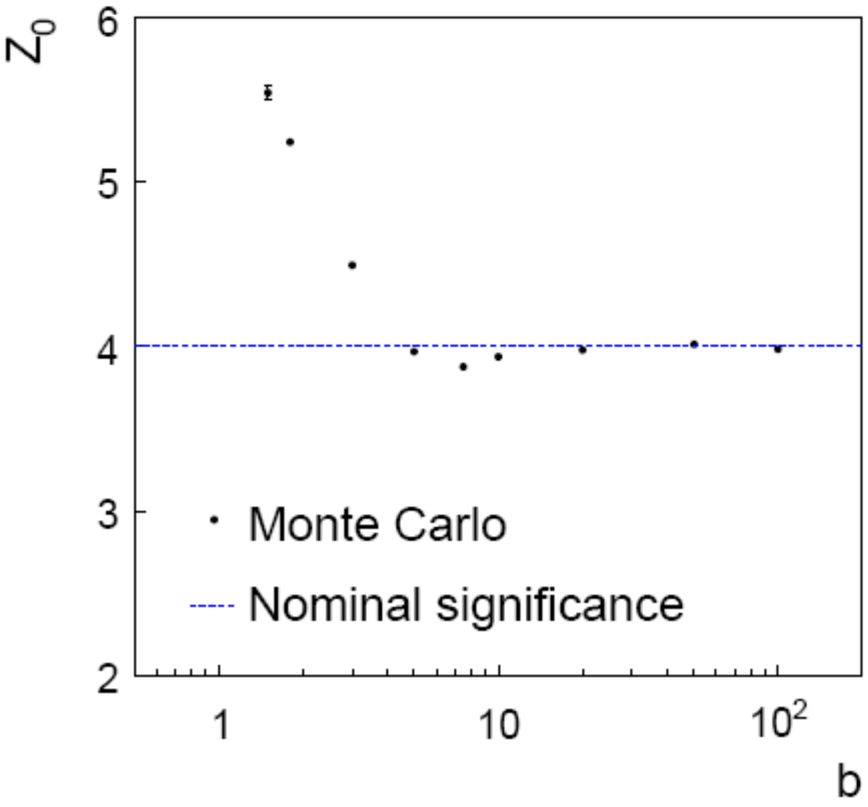
$$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0, \\ \dfrac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0 \; . \end{cases} \qquad \tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

Performance not identical to but very close to q_μ (of previous slide).

q_μ is simpler in important ways: asymptotic distribution is independent of nuisance parameters.

Monte Carlo test of asymptotic formulae

Significance from asymptotic formula, here $Z_0 = \sqrt{q_0} = 4$, compared to MC (true) value.



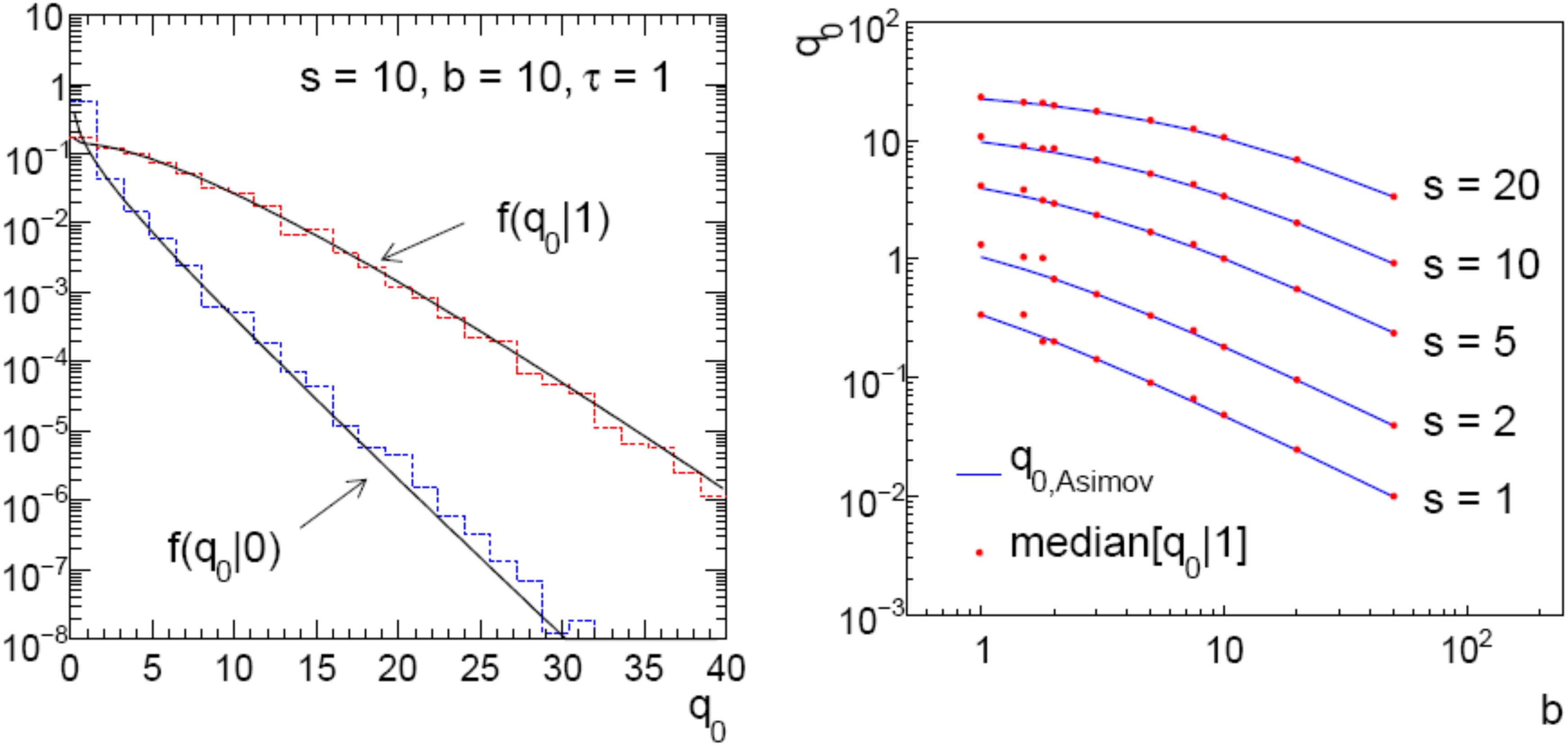
For very low *b*, asymptotic formula underestimates Z_0.

Then slight overshoot before rapidly converging to MC value.

Monte Carlo test of asymptotic formulae

Asymptotic $f(q_0|1)$ good already for fairly small samples.

Median$[q_0|1]$ from Asimov data set; good agreement with MC.



Feldman-Cousins discussion

The initial motivation for Feldman-Cousins (unified) confidence intervals was to eliminate null intervals.

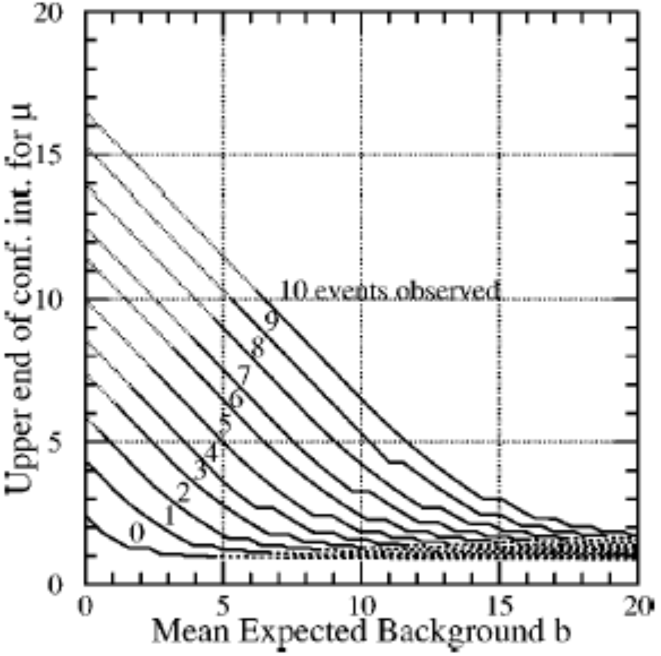
The F-C limits are based on a likelihood ratio for a test of μ with respect to the alternative consisting of all other allowed values of μ (not just, say, lower values).

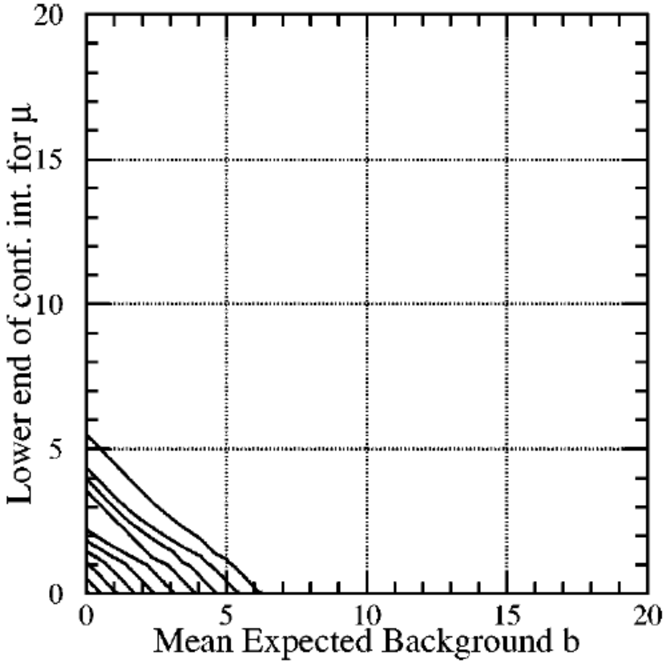
The interval's upper edge is higher than the limit from the one-sided test, and lower values of μ may be excluded as well. A substantial downward fluctuation in the data gives a low (but nonzero) limit.

This means that when a value of μ is excluded, it is because there is a probability α for the data to fluctuate either high or low in a manner corresponding to less compatibility as measured by the likelihood ratio.

Upper/lower edges of F-C interval for μ versus b for $n \sim \text{Poisson}(\mu+b)$

Feldman & Cousins, PRD 57 (1998) 3873





Lower edge may be at zero, depending on data.

For $n = 0$, upper edge has (weak) dependence on b.