Recent developments in statistical methods for particle physics

BERGISCHE
UNIVERSITÄT
WUPPERTAL

Particle Physics Seminar
Wuppertal, 14 July 2011

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

Outline

Developments related to setting limits (CLs, PCL, F-C, etc.)

CCGV arXiv:1105.3166

Asymptotic formulae for distributions of test statistics based on the profile likelihood ratio

CCGV, arXiv:1007.1727, EPJC 71 (2011) 1-19

Other recent developments

The Look-Elsewhere Effect, Gross and Vitells, arXiv:1005.1891, Eur.Phys.J.C70:525-530,2010

Reminder about statistical tests

Consider test of a parameter μ, e.g., proportional to cross section.

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

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

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

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

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

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

Test statistics and p-values

Often construct a test statistic, q_μ, which reflects the level of agreement between the data and the hypothesized value μ.

For examples of statistics based on the profile likelihood ratio, see, e.g., CCGV arXiv:1007.1727 (the "Asimov" paper).

Usually define q_μ such that higher values represent increasing incompatibility with the data, so that the p-value of μ is:

$$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu)\, dq_\mu$$

observed value of q_μ

pdf of q_μ assuming μ

Equivalent formulation of test: reject μ if $p_\mu < \alpha$.

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 confidence interval will by construction contain the true value of μ with probability of at least $1 - \alpha$.

Can give upper limit μ_{up}, i.e., the largest value of μ not rejected, i.e., the upper edge of the confidence interval.

The interval (and limit) depend on the choice of the test, which is often based on considerations of power.

Power of a statistical test

But where to define critical region? Usually put this where the test has a high *power* with respect to an alternative hypothesis μ'.

The *power* of the test of μ with respect to the alternative μ' is the probability to reject μ if μ' is true:

(M = Mächtigkeit, мощность)

$$M_{\mu'}(\mu) = P(\mathbf{x} \in w_\mu | \mu') = P(p_\mu < \alpha | \mu')$$

E.g., for an upper limit, maximize the power with respect to the alternative consisting of $\mu' < \mu$.

Other types of tests not based directly on power (e.g., likelihood ratio).

Choice of test for limits

Often 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.

To do this take the alternative to correspond to lower values of μ.

The critical region to test μ thus contains low values of the data.

→ One-sided (e.g., upper) limit.

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 (e.g., likelihood ratio wrt two-sided alternative).

The critical region can contain both high and low data values.

→ Two-sided or unified (Feldman-Cousins) intervals.

Test statistic for upper limits

For purposes of setting an upper limit on μ use

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

I.e. 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: $$p_\mu = \int_{q_{\mu,\mathrm{obs}}}^{\infty} f(q_\mu|\mu)\, dq_\mu$$

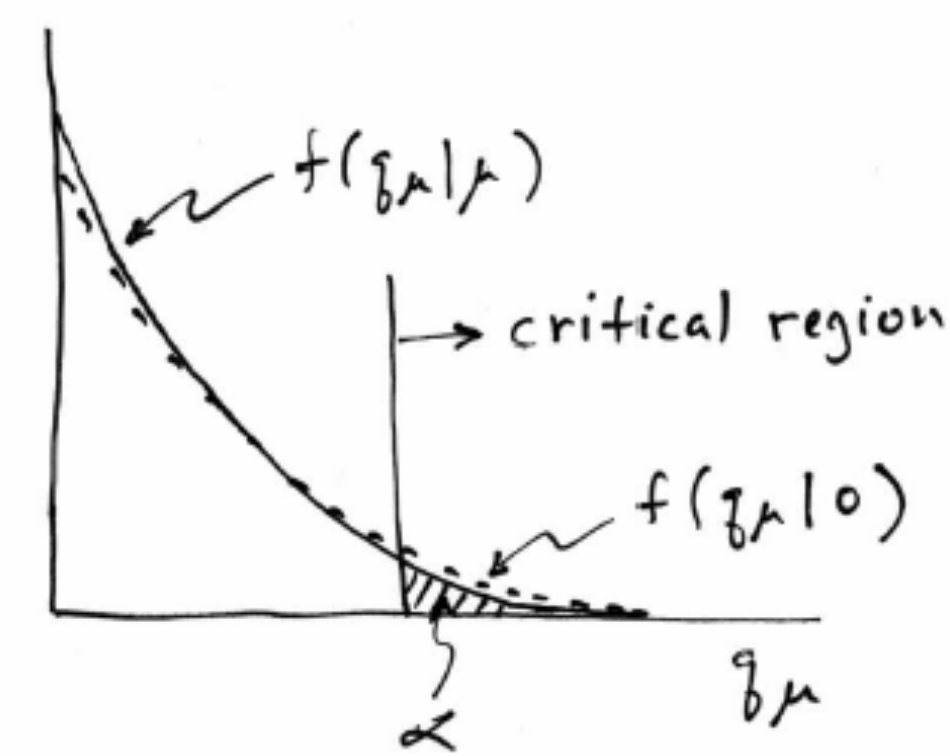
Large sample approximation: $$p_\mu = 1 - \Phi\left(\sqrt{q_\mu}\right)$$

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

Low sensitivity to μ

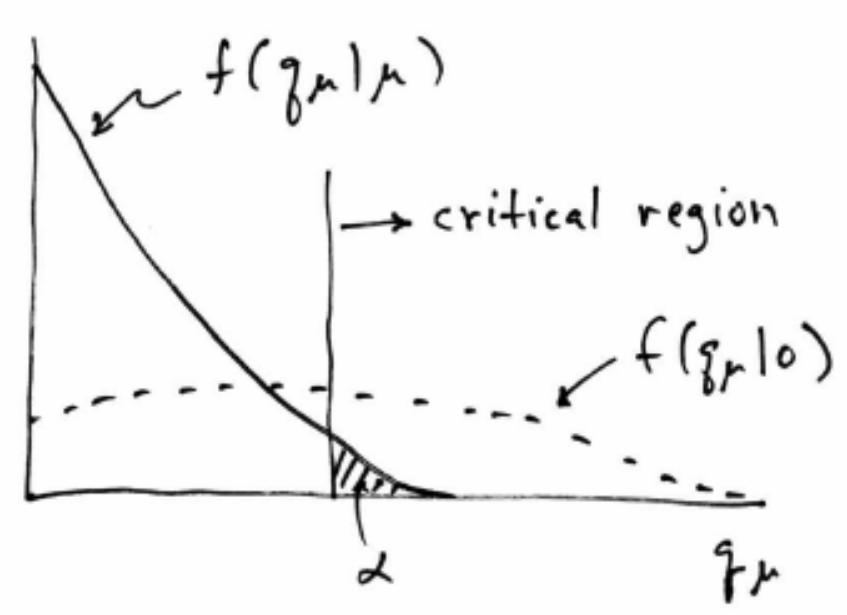
It can be that the effect of a given hypothesized μ is very small relative to the background-only ($\mu = 0$) prediction.

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:

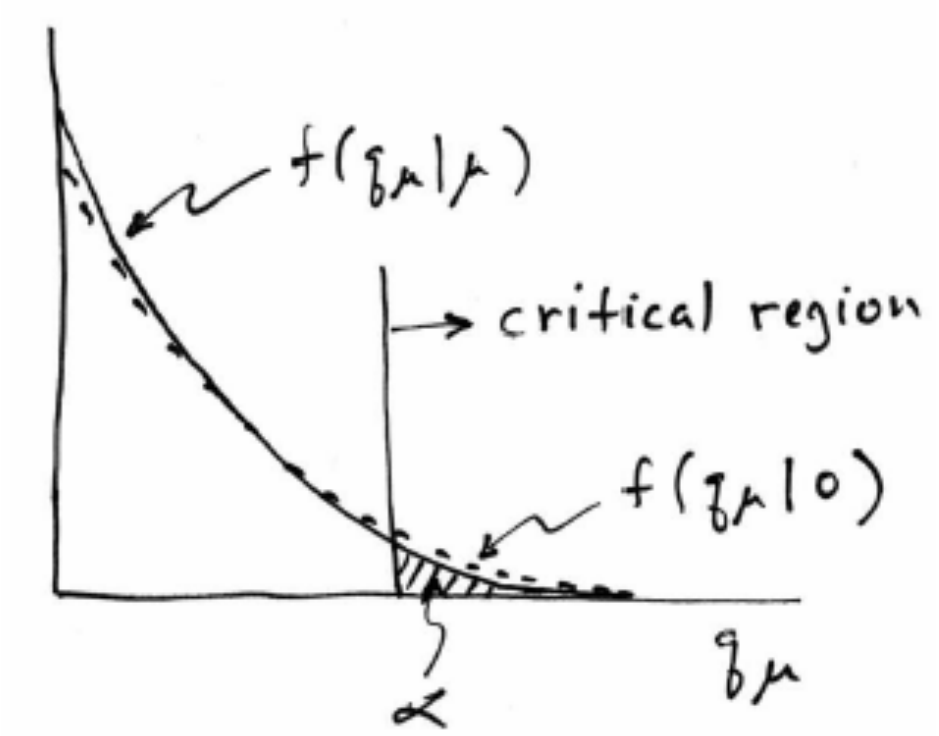


That is, the power (probability to reject μ if $\mu = 0$) is substantially higher than α. We use this power as a measure of the sensitivity.

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., $m_{\mathbf{H}} = 1000$ TeV).

"Spurious exclusion"

Previous 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.

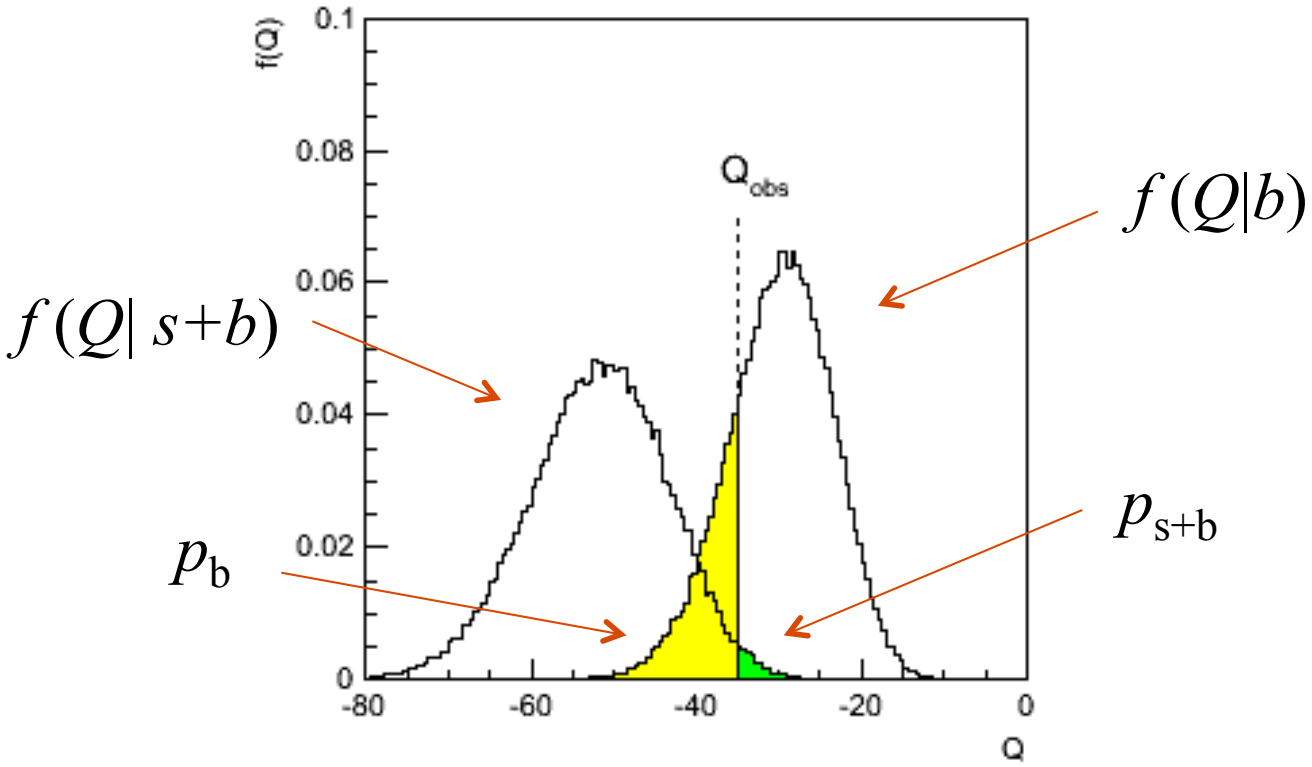
In the 1990s this was re-examined for the LEP Higgs search by Alex Read and others

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.

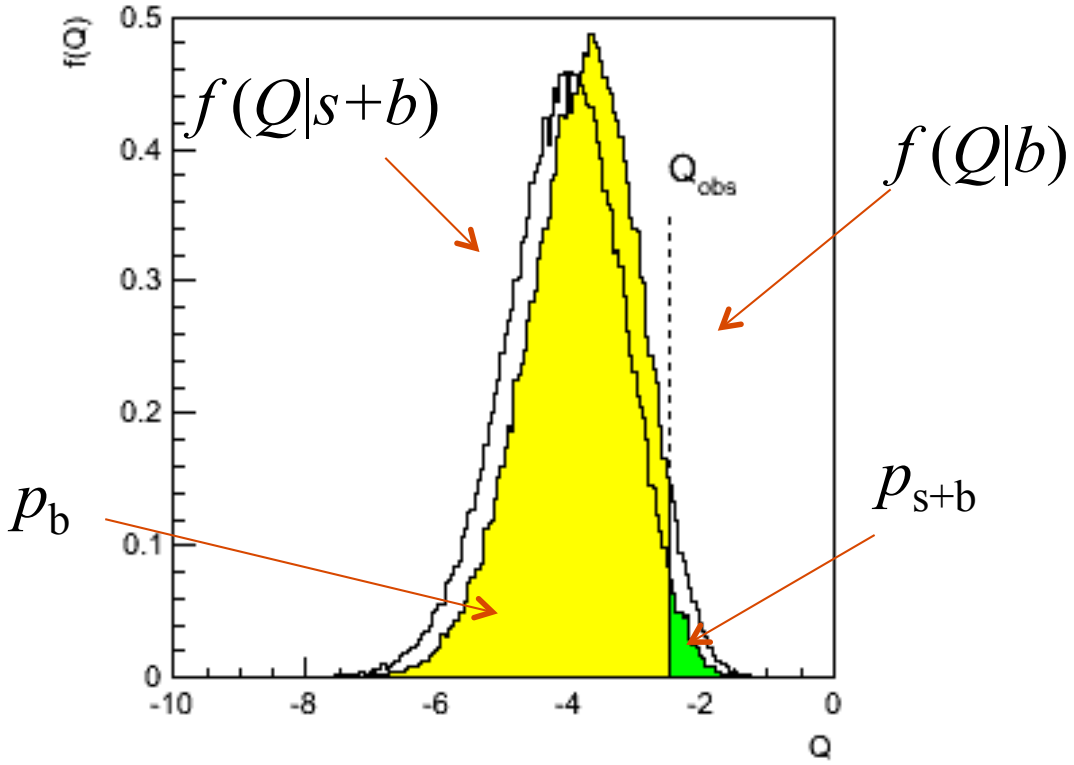
The CL_s procedure

In the usual formulation of CL_s, one tests both the $\mu = 0$ (b) and $\mu = 1$ (s+b) hypotheses with the same statistic $Q = -2\ln L_{s+b}/L_b$:



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_s solution (A. Read et al.) is to base the test not on the usual *p*-value (CL_{s+b}), but rather to divide this by CL_b (one minus the *p*-value of the *b*-only hypothesis, i.e.,

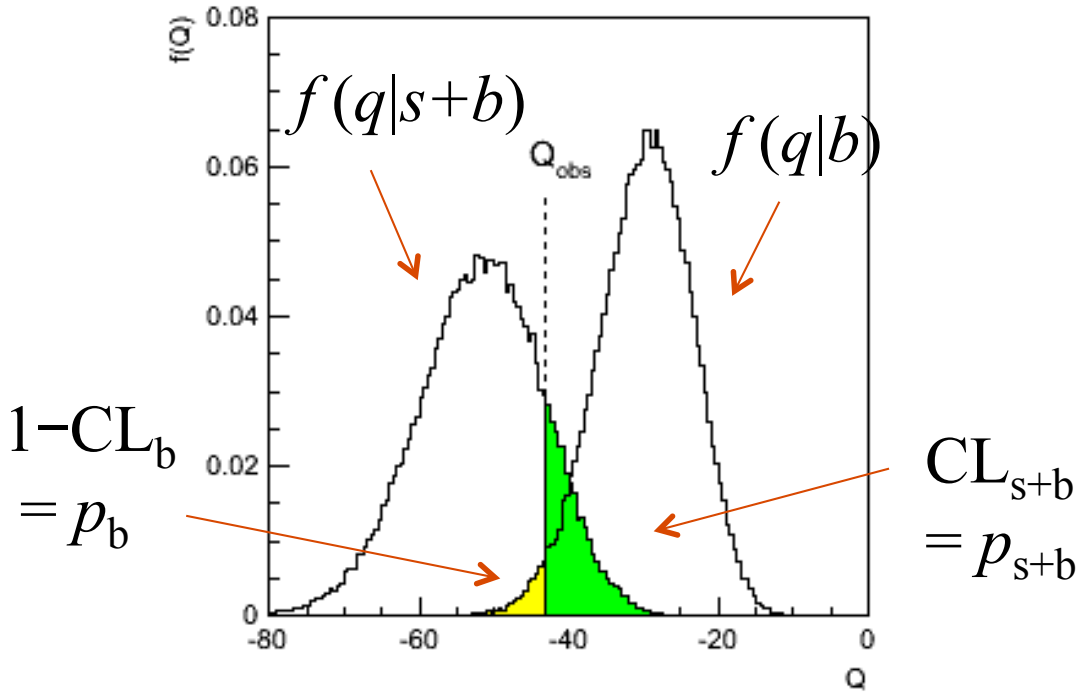
Define:

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

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

Reject s+b hypothesis if:

$$CL_s \leq \alpha$$



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

Feldman-Cousins unified intervals

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.

Power Constrained Limits (PCL)

CL_s has been criticized because the coverage probability of the upper limit is greater than the nominal CL = 1 − α by an amount that is not readily apparent (but can be computed).

Therefore we have proposed an alternative method for protecting against exclusion with little/no sensitivity, by regarding a value of μ to be excluded if:

(a) the value μ is rejected by the test, i.e., $\mathbf{x} \in w_\mu$ or equivalently $p_\mu < \alpha$, and

(b) one has sufficient sensitivity to μ, i.e., $M_0(\mu) \geq M_{\text{min}}$.

Here the measure of sensitivity is the power of the test of μ with respect to the alternative $\mu = 0$:

$$M_0(\mu) = P(\mathbf{x} \in w_\mu | 0) = P(p_\mu < \alpha | 0)$$

Constructing PCL

First compute the distribution under assumption of the background-only ($\mu = 0$) hypothesis of the "usual" upper limit $\mu_{\rm up}$ with no power constraint.

The power of a test of μ with respect to $\mu = 0$ is the fraction of times that μ is excluded ($\mu_{\rm up} < \mu$):

$$M_0(\mu) = P(\mu_{\rm up} < \mu|0)$$

Find the smallest value of μ ($\mu_{\rm min}$), such that the power is at least equal to the threshold $M_{\rm min}$.

The Power-Constrained Limit is:

$$\mu^*_{\rm up} = \max(\mu_{\rm up}, \mu_{\rm min})$$

PCL for upper limit with Gaussian measurement

Suppose $\hat{\mu} \sim \text{Gauss}(\mu, \sigma)$, goal is to set upper limit on μ.

Define critical region for test of μ as $\quad \hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha)$

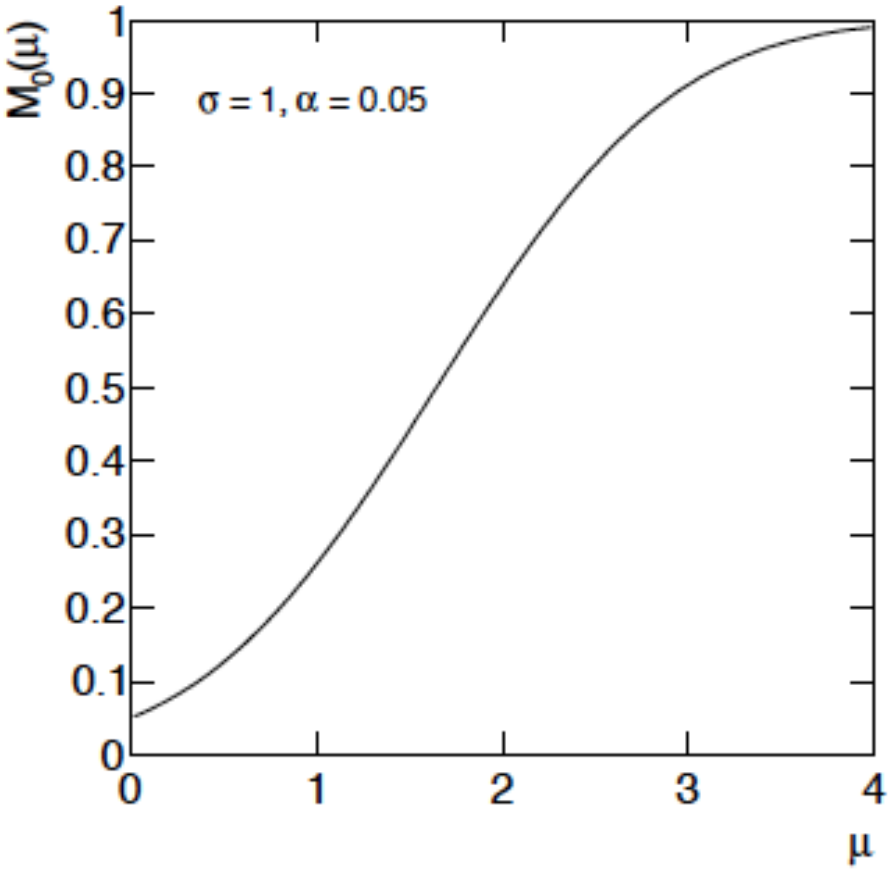
inverse of standard Gaussian cumulative distribution

This gives (unconstrained) upper limit: $\quad \mu_{\text{up}} = \hat{\mu} + \sigma\Phi^{-1}(1-\alpha)$

Power $M_0(\mu)$ for Gaussian measurement

The power of the test of μ with respect to the alternative $\mu' = 0$ is:

$$M_0(\mu) = P\left(\hat{\mu} < \mu - \sigma\Phi^{-1}(1-\alpha)|0\right) = \Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right)$$



standard Gaussian cumulative distribution

Spurious exclusion when $\hat{\mu}$ fluctuates down

Requiring the power be at least M_{min}

$$\Phi\left(\frac{\mu}{\sigma} - \Phi^{-1}(1-\alpha)\right) \geq M_{\text{min}}$$

implies that the smallest μ to which one is sensitive is

$$\mu_{\text{min}} = \sigma\left(\Phi^{-1}(M_{\text{min}}) + \Phi^{-1}(1-\alpha)\right)$$

If one were to use the unconstrained limit, values of μ at or below μ_{min} would be excluded if

$$\hat{\mu} < \sigma\Phi^{-1}(M_{\text{min}})$$

That is, one excludes $\mu < \mu_{\text{min}}$ when the unconstrained limit fluctuates too far downward.

Choice of minimum power

Choice of M_{min} is convention. Formally it should be large relative to α (5%). Earlier we have proposed

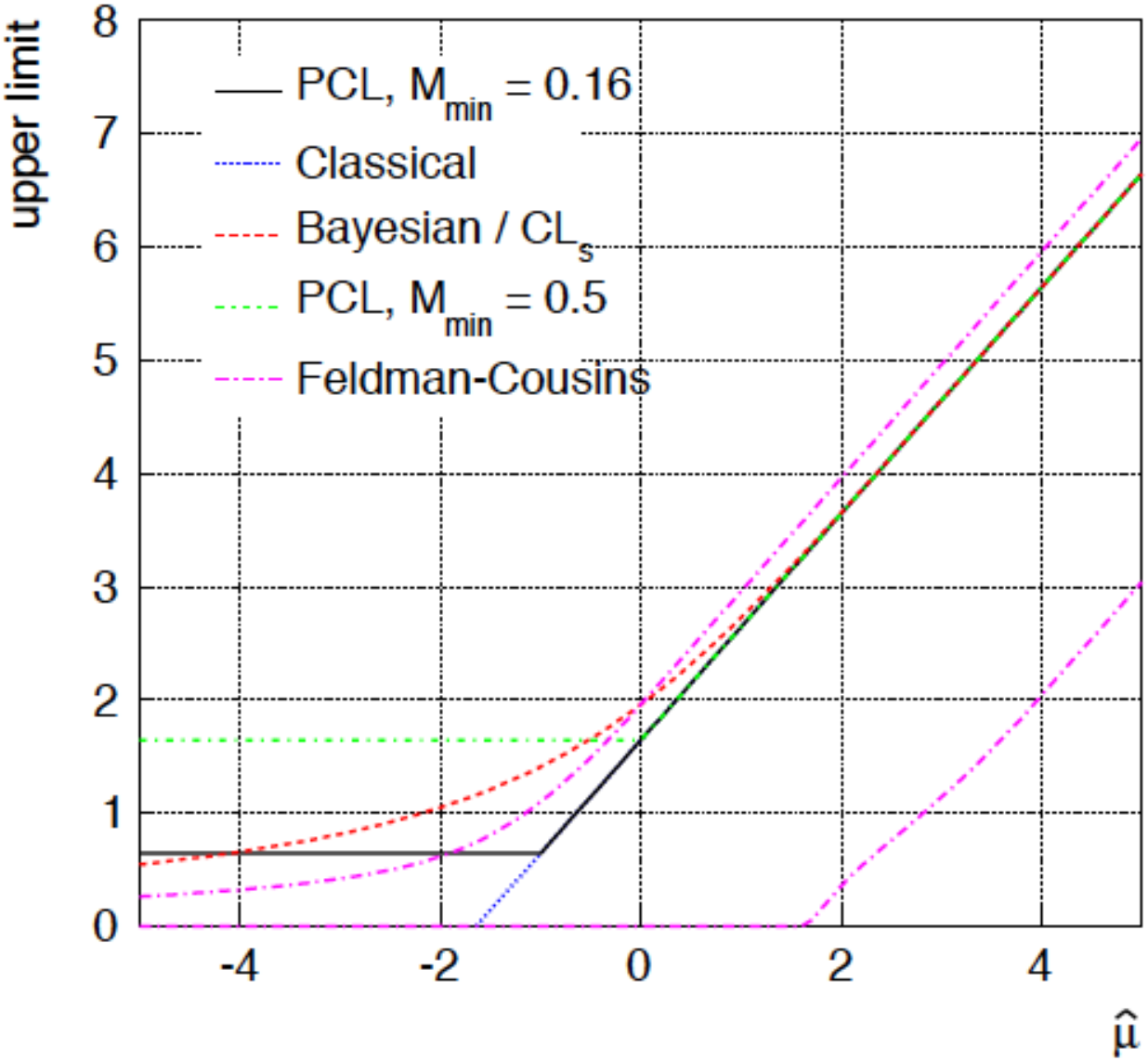
$$M_{\text{min}} = \Phi(-1) = 0.1587$$

because in Gaussian example this means that one applies the power constraint if the observed limit fluctuates down by one standard deviation.

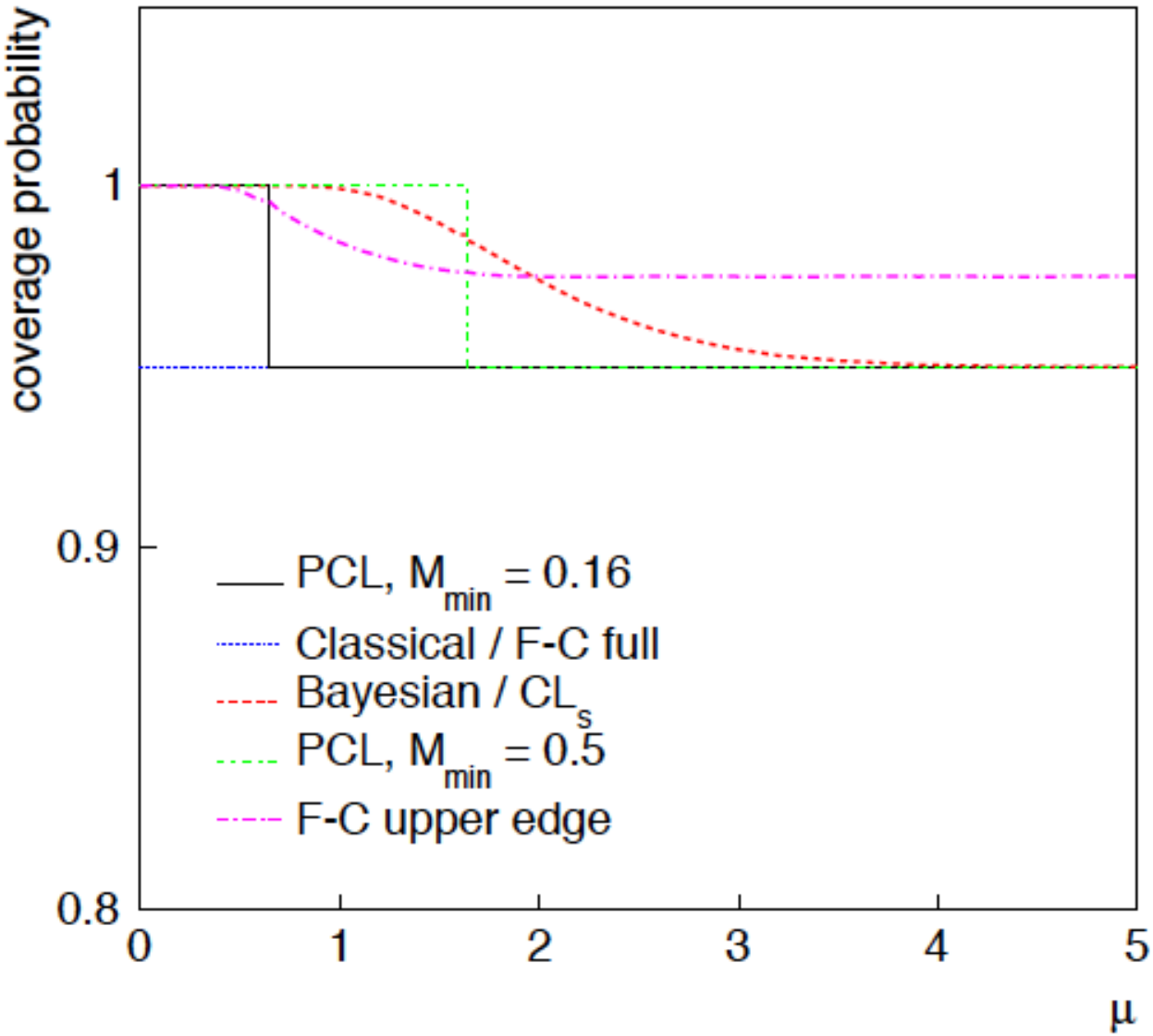
In fact the distribution of μ_{up} is often roughly Gaussian, so we call this a "1σ" (downward) fluctuation and use $M_{\text{min}} = 0.16$ regardless of the exact distribution of μ_{up}.

For the Gaussian example, this gives $\mu_{\text{min}} = 0.64\sigma$, i.e., the lowest limit is similar to the intrinsic resolution of the measurement (σ).

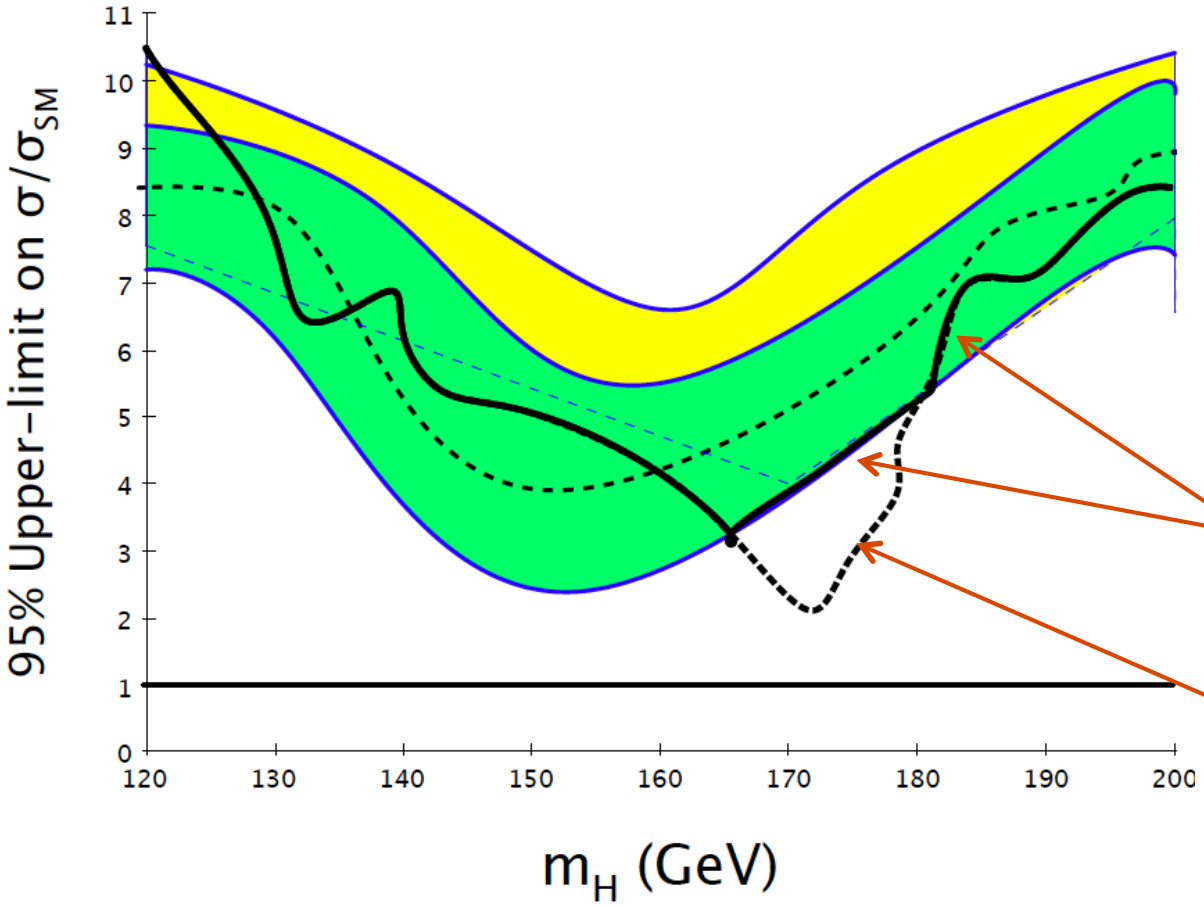
Upper limits for Gaussian problem



Coverage probability for Gaussian problem



PCL as a function of, e.g., m_H



PCL

Here power below threshold; do not exclude.

Some reasons to consider increasing $M_{\min}$

$M_{\min}$ is supposed to be "substantially" greater than α (5%).

So $M_{\min} = 16\%$ is fine for $1 - \alpha = 95\%$, but if we ever want $1 - \alpha = 90\%$, then16% is not "large" compared to 10%; $\mu_{\min} = 0.28\sigma$ starts to look small relative to the intrinsic resolution of the measurement. Not an issue if we stick to 95% CL.

PCL with $M_{\min} = 16\%$ is often substantially lower than CLs. This is because of the conservatism of CLs (see coverage).

But goal is not to get a lower limit per se, rather

- to use a test with higher power in those regions where one feels there is enough sensitivity to justify exclusion and
- to allow for easy communication of coverage (95% for $\mu \geq \mu_{\min}$; 100% otherwise).

A few further considerations

It could be that owing to practical constraints, certain systematic uncertainties are over-estimated in an analysis; this could be justified by wanting to be conservative. This means the +/−1 sigma bands of the unconstrained limit are broader (but the median should move up), and it could happen that the PCL limit for $M_{\text{min}} = 16\%$ becomes lower (conservative = aggresive).

Obtaining PCL requires the distribution of unconstrained limits, from which one finds the M_{min} (16%, 50%) percentile.

In some analyses this can entail calculational issues that are expected to be less problematic for $M_{\text{min}} = 50\%$ than for 16%.

Analysts produce anyway the median limit, even in absence of the error bands, so with $M_{\text{min}} = 50\%$ the burden on the analyst is reduced somewhat (but one would still want the error bands).

We therefore recently proposed moving M_{min} to 50%.

Treatment of nuisance parameters

In most problems, the data distribution is not uniquely specified by μ but contains nuisance parameters θ.

This makes it more difficult to construct an (unconstrained) interval with correct coverage probability for all values of θ, so sometimes approximate methods used ("profile construction").

More importantly for PCL, the power $M_0(\mu)$ can depend on θ. So which value of θ to use to define the power?

Since the power represents the probability to reject μ if the true value is $\mu = 0$, to find the distribution of μ_{up} we take the values of θ that best agree with the data for $\mu = 0$: $\hat{\hat{\boldsymbol{\theta}}}(0)$

May seem counterintuitive, since the measure of sensitivity now depends on the data. We are simply using the data to choose the most appropriate value of θ where we quote the power.

Summary on CLs, PCL, etc.

With a "usual" confidence limit, a large downward fluctuation can lead to exclusion of parameter values to which one has little or no sensitivity (will happen 5% of the time).

PCL solves this problem by separating the parameter space into regions within which one has/hasn't sufficient sensitivity as given by the probability to reject μ if background-only model is true.

Current recommendation: power $M_0(\mu) \geq 0.5$.

Within region with sufficient sensitivity, an upper limit can be set with a one-sided test (highest power) and exact $1 - \alpha$ coverage.

It is important to report both the constrained and unconstrained limits, so one can see where the power constraint comes into play.

Procedure easily adapted to problems with nuisance parameters (quote power at estimated values of nuisance parameters for $\mu = 0$).

More recent developments

Large-sample statistical formulae for a search at the LHC

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727,
EPJC 71 (2011) 1-19

Significance test using profile likelihood ratio

Systematics included via nuisance parameters

Distributions in large sample limit, no MC used.

Progress on related issues (some updates from PHYSTAT2011):

The "look elsewhere effect"

Combining measurements (RooStats)

Prototype search analysis

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

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

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 ($\boldsymbol{\theta}_s$, $\boldsymbol{\theta}_b$,b_{tot})

Likelihood function is

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

The profile likelihood ratio

Base significance test on the profile likelihood ratio:

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

maximizes L for Specified μ

maximize L

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

The profile LR hould be near-optimal in present analysis with variable μ and nuisance parameters $\boldsymbol{\theta}$.

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.

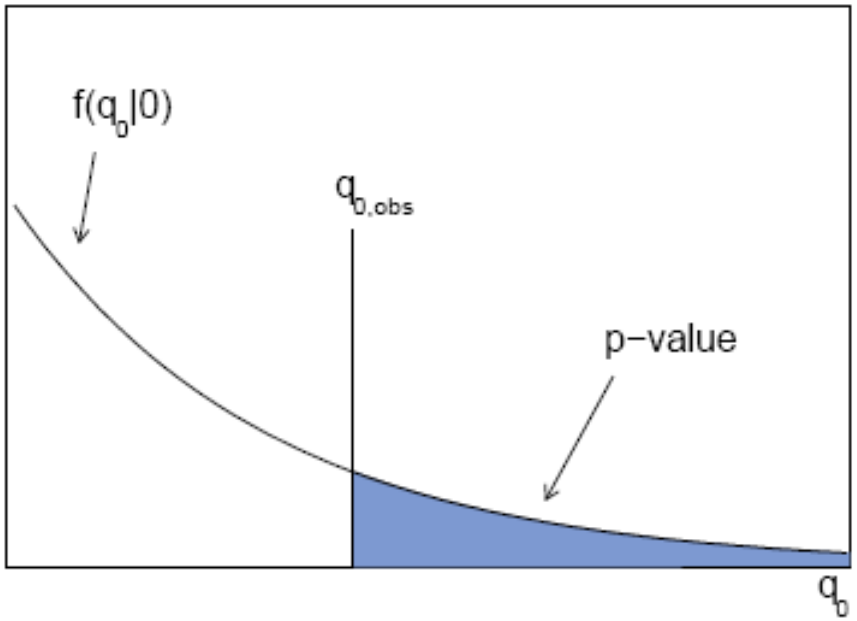
Note that even though here physically $\mu \geq 0$, we allow $\hat{\mu}$ to be negative. In large sample limit its distribution becomes Gaussian, and this will allow us to write down simple expressions for distributions of our test statistics.

p-value for discovery

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

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

will get formula for this later

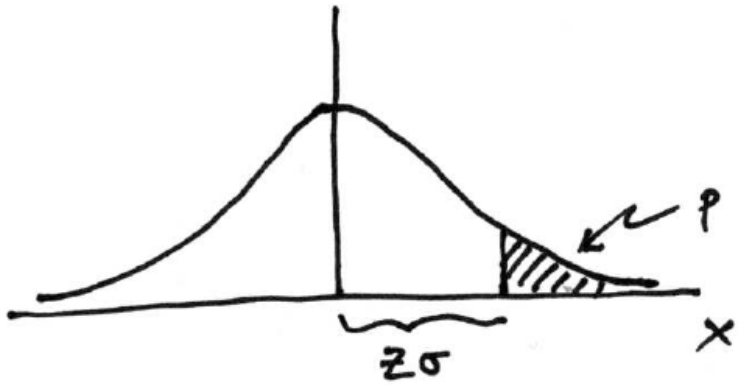


From *p*-value get equivalent significance,

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

Significance from p-value

Often define significance Z as the number of standard deviations that a Gaussian variable would fluctuate in one direction to give the same p-value.

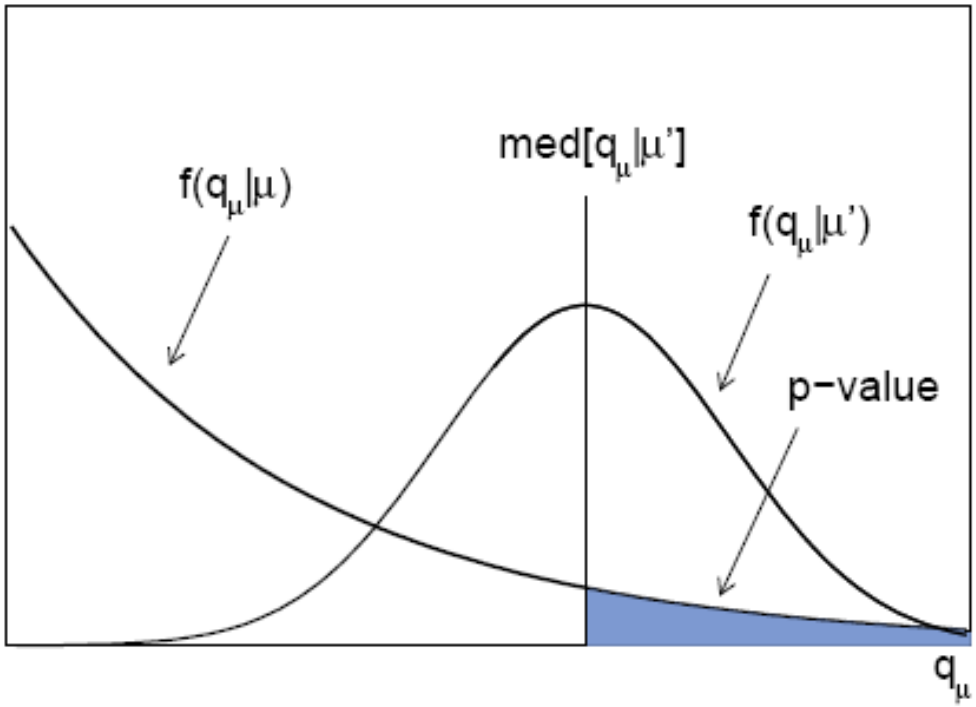


$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$ `1 - TMath::Freq`

$$Z = \Phi^{-1}(1-p)$$ `TMath::NormQuantile`

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 $f(q_0|\mu')$,

Test statistic for upper limits

For purposes of setting an upper limit on μ 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.

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.

Wald approximation for profile likelihood ratio

To find *p*-values, we need: $f(q_0|0), \quad f(q_\mu|\mu)$

For median significance under alternative, need: $f(q_\mu|\mu')$

Use approximation due to Wald (1943)

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

$$\hat{\mu} \sim \text{ Gaussian}(\mu', \sigma)$$

sample size

i.e., $E[\hat{\mu}] = \mu'$

σ from covariance matrix V, use, e.g.,

$$V^{-1} = -E\left[\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right]$$

Noncentral chi-square for $-2\ln\lambda(\mu)$

If we can neglect the $O(1/\sqrt{N})$ term, $-2\ln\lambda(\mu)$ follows a noncentral chi-square distribution for one degree of freedom with noncentrality parameter

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}$$

As a special case, if $\mu' = \mu$ then $\Lambda = 0$ and $-2\ln\lambda(\mu)$ follows a chi-square distribution for one degree of freedom (Wilks).

The Asimov data set

To estimate median value of $-2\ln\lambda(\mu)$, consider special data set where all statistical fluctuations suppressed and n_i, m_i are replaced by their expectation values (the "Asimov" data set):

$$n_i = \mu' s_i + b_i$$

$$m_i = u_i$$

$$\rightarrow \quad \hat{\mu} = \mu' \qquad \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$$

$$\lambda_{\rm A}(\mu) = \frac{L_{\rm A}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_{\rm A}(\hat{\mu}, \hat{\boldsymbol{\theta}})} = \frac{L_{\rm A}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_{\rm A}(\mu', \boldsymbol{\theta})}$$

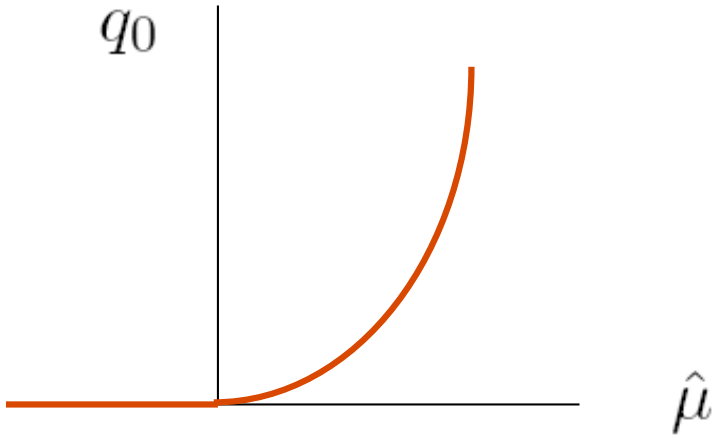
$$-2\ln\lambda_{\rm A}(\mu) = \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda$$

Asimov value of $-2\ln\lambda(\mu)$ gives non-centrality param. Λ, or equivalently, σ.

Relation between test statistics and $\hat{\mu}$

Assuming Wald approximation, the relation between q_0 and $\hat{\mu}$ is

$$q_0 = \begin{cases} \hat{\mu}^2/\sigma^2 & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$



Monotonic, therefore quantiles of $\hat{\mu}$ map one-to-one onto those of q_0, e.g.,

$$\text{med}[q_0] = q_0(\text{med}[\hat{\mu}]) = q_0(\mu') = \frac{\mu'^2}{\sigma^2} = -2\ln\lambda_A(0)$$

Distribution of q_0

Assuming the Wald approximation, 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:

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

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

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

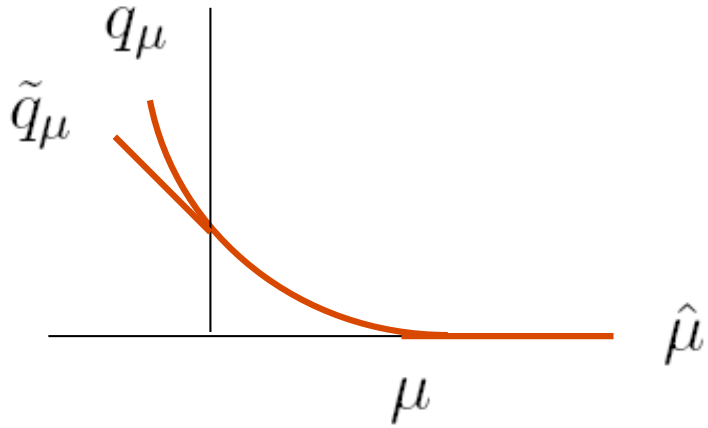
Therefore the discovery significance Z is simply

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

Relation between test statistics and $\hat{\mu}$

Assuming the Wald approximation for $-2\ln\lambda(\mu)$, q_μ and $\tilde{q}_\mu$ both have monotonic relation with μ.

$$q_\mu = \begin{cases} \frac{(\mu-\hat{\mu})^2}{\sigma^2} & \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$



$$\tilde{q}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0 \\ \frac{(\mu-\hat{\mu})^2}{\sigma^2} & 0 \le \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \,, \end{cases}$$

And therefore quantiles of q_μ, $\tilde{q}_\mu$ can be obtained directly from those of $\hat{\mu}$ (which is Gaussian).

Distribution of q_μ

Similar results for q_μ

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

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

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

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

Distribution of $\tilde{q}_\mu$

Similar results for $\tilde{q}_\mu$

$$f(\tilde{q}_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\tilde{q}_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{\tilde{q}_\mu}-\frac{\mu-\mu'}{\sigma}\right)^2\right] & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2\ , \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)}\exp\left[-\frac{1}{2}\frac{(\tilde{q}_\mu-(\mu^2-2\mu\mu')/\sigma^2)^2}{(2\mu/\sigma)^2}\right] & \tilde{q}_\mu > \mu^2/\sigma^2\ . \end{cases}$$

$$F(\tilde{q}_\mu|\mu') = \begin{cases} \Phi\left(\sqrt{\tilde{q}_\mu}-\frac{(\mu-\mu')}{\sigma}\right) & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2\ , \\ \Phi\left(\frac{\tilde{q}_\mu-(\mu^2-2\mu\mu')/\sigma^2}{2\mu/\sigma}\right) & \tilde{q}_\mu > \mu^2/\sigma^2\ . \end{cases}$$

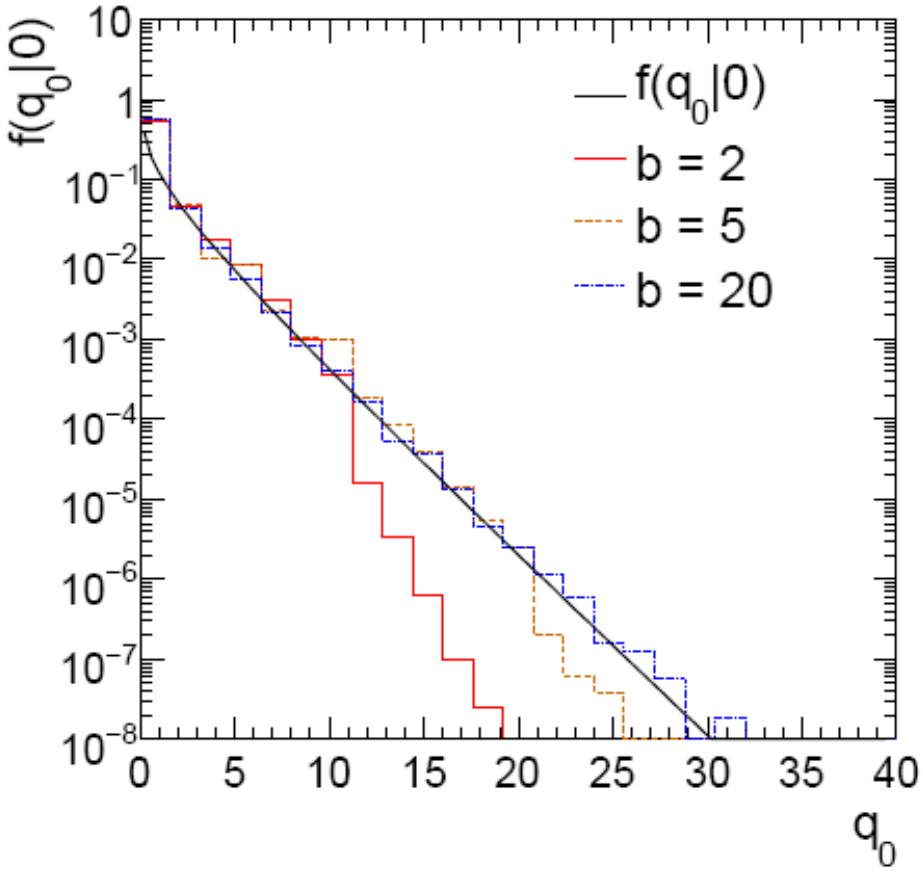
Monte Carlo test of asymptotic formula

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

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

Here take $\tau = 1$.

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

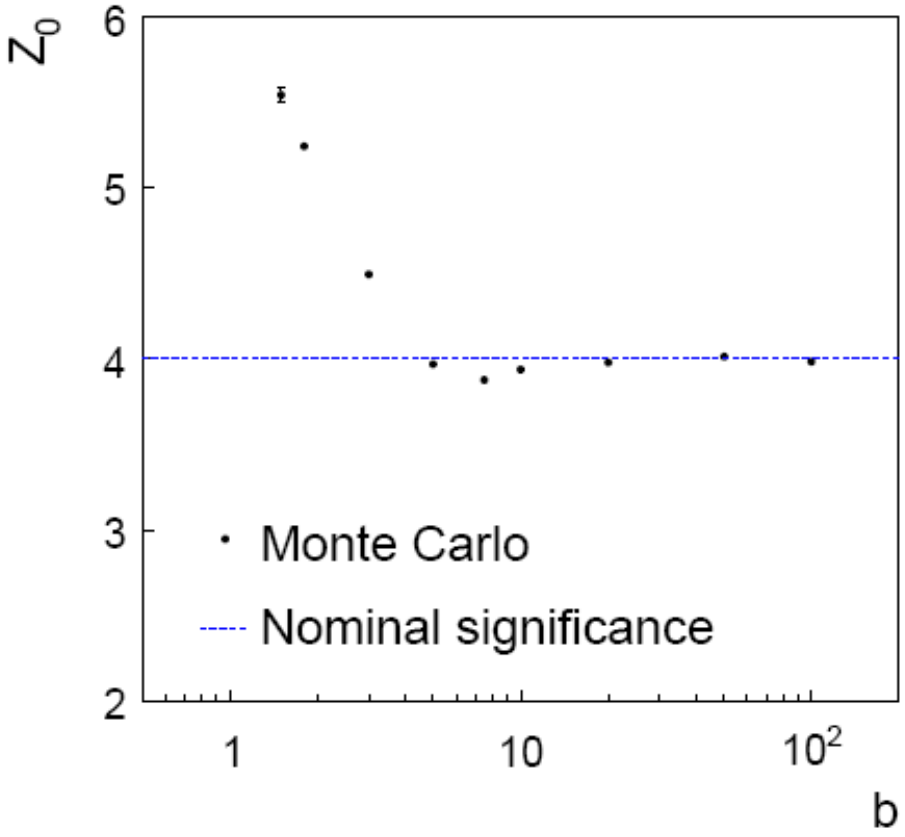


Monte Carlo test of asymptotic formulae

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.

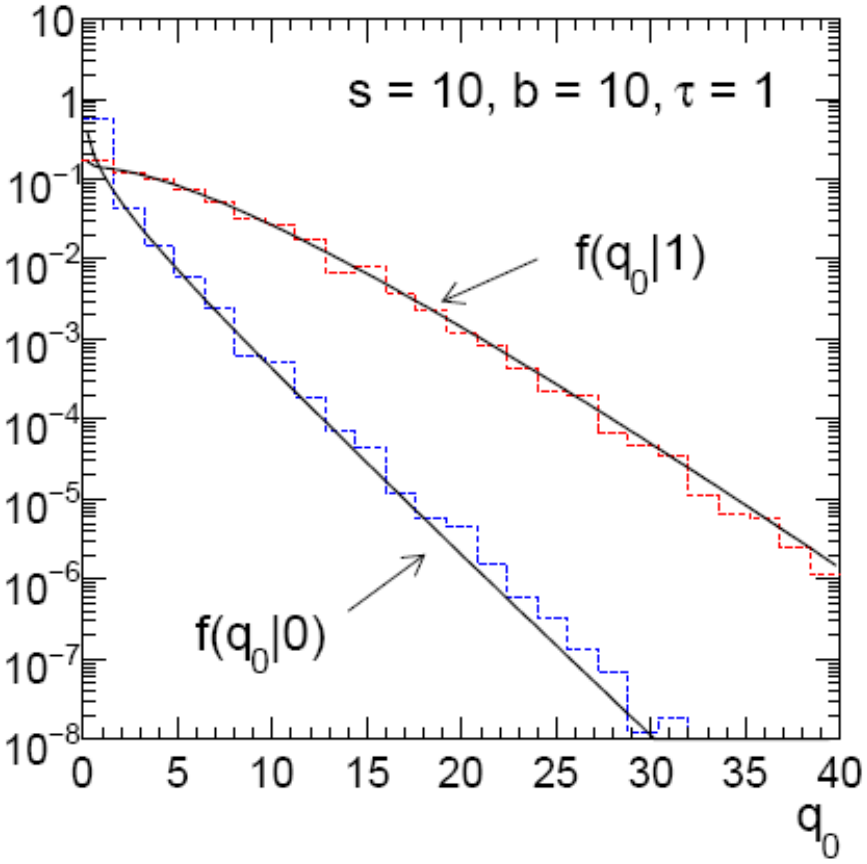
Then slight overshoot before rapidly converging to MC value.

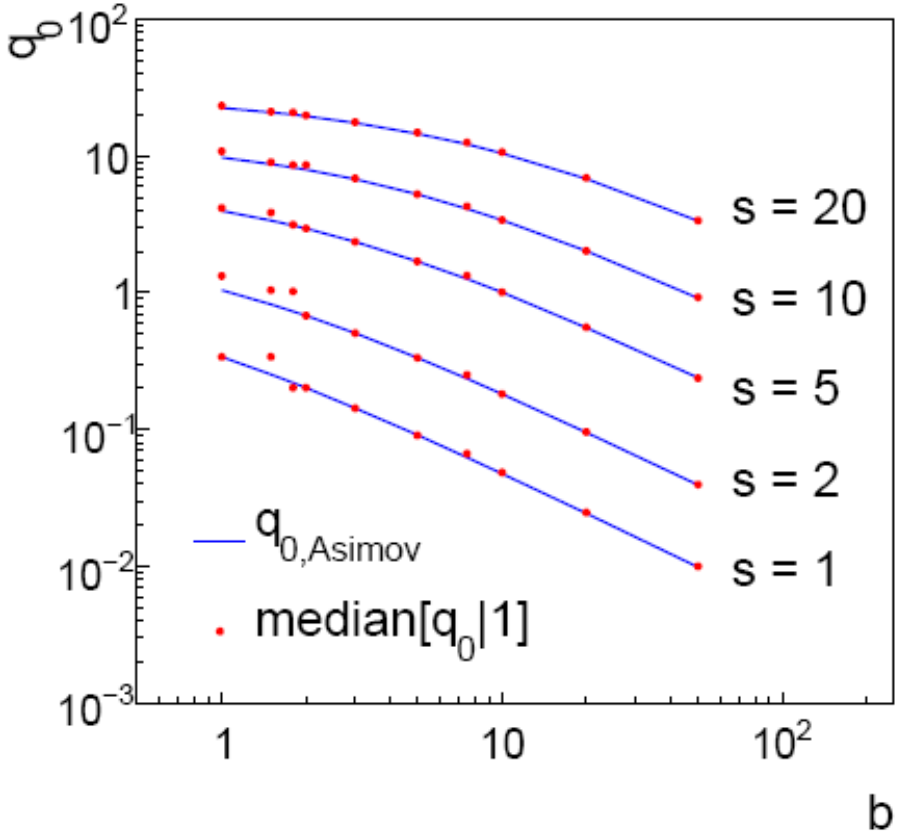


Monte Carlo test of asymptotic formulae

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

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





Monte Carlo test of asymptotic formulae

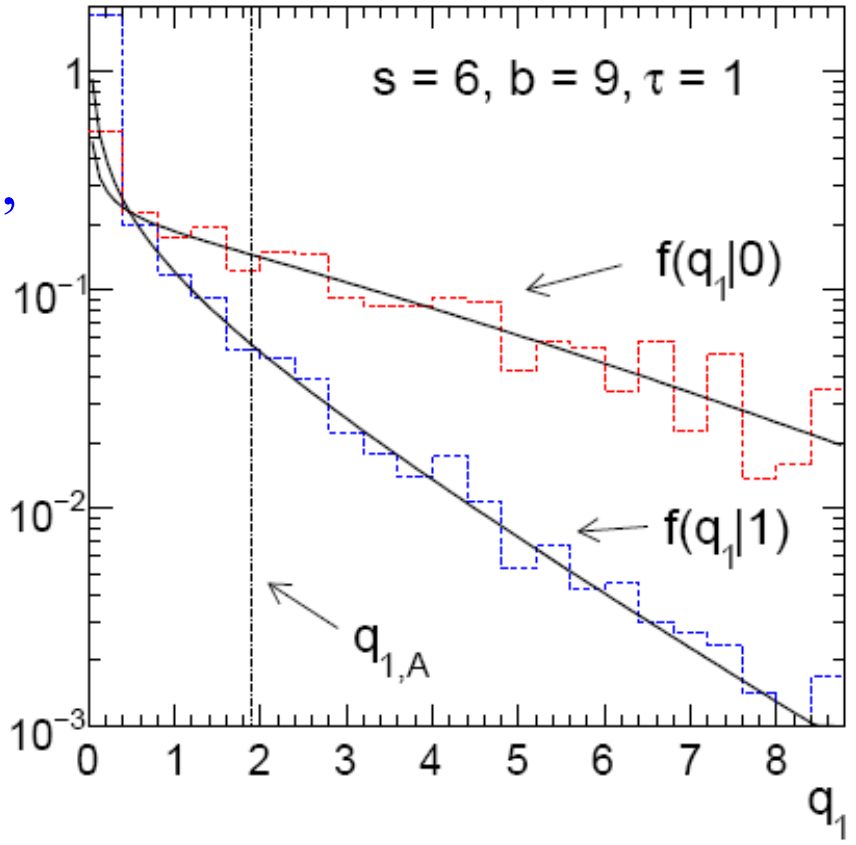
Consider again $n \sim \text{Poisson}(\mu s + b)$, $m \sim \text{Poisson}(\tau b)$
Use q_μ to find p-value of hypothesized μ values.

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".

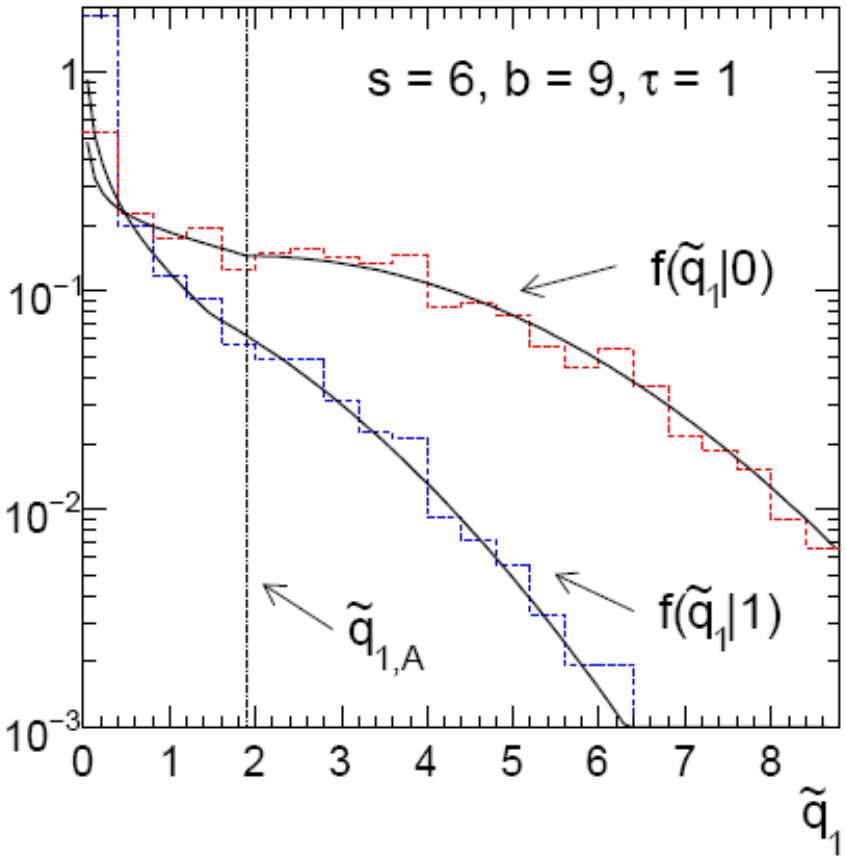
Here asymptotic formulae good for $s = 6$, $b = 9$.



Monte Carlo test of asymptotic formulae

Same message for test based on $\tilde{q}_\mu$.

q_μ and $\tilde{q}_\mu$ give similar tests to the extent that asymptotic formulae are valid.



Discovery significance for $n \sim \text{Poisson}(s + b)$

Consider again the case where we observe n events , model as following Poisson distribution with mean $s + b$ (assume b is known).

1) For an observed n, what is the significance Z_0 with which we would reject the $s = 0$ hypothesis?

2) What is the expected (or more precisely, median) Z_0 if the true value of the signal rate is s?

Gaussian approximation for Poisson significance

For large $s + b$, $n \to x \sim$ Gaussian(μ,σ) , $\mu = s + b$, $\sigma = \sqrt{(s + b)}$.

For observed value $x_{\rm obs}$, p-value of $s = 0$ is Prob($x > x_{\rm obs} \mid s = 0$),:

$$p_0 = 1 - \Phi\left(\frac{x_{\rm obs} - b}{\sqrt{b}}\right)$$

Significance for rejecting $s = 0$ is therefore

$$Z_0 = \Phi^{-1}(1 - p_0) = \frac{x_{\rm obs} - b}{\sqrt{b}}$$

Expected (median) significance assuming signal rate s is

$$\text{median}[Z_0|s + b] = \frac{s}{\sqrt{b}}$$

Better approximation for Poisson significance

Likelihood function for parameter s is

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

or equivalently the log-likelihood is

$$\ln L(s) = n \ln(s+b) - (s+b) - \ln n!$$

Find the maximum by setting $\dfrac{\partial \ln L}{\partial s} = 0$

gives the estimator for s: $\hat{s} = n - b$

Approximate Poisson significance (continued)

The likelihood ratio statistic for testing $s = 0$ is

$$q_0 = -2 \ln \frac{L(0)}{L(\hat{s})} = 2 \left(n \ln \frac{n}{b} + b - n \right) \quad \text{for } n > b, \ 0 \text{ otherwise}$$

For sufficiently large $s + b$, (use Wilks' theorem),

$$Z_0 \approx \sqrt{q_0} = \sqrt{2 \left(n \ln \frac{n}{b} + b - n \right)} \quad \text{for } n > b, \ 0 \text{ otherwise}$$

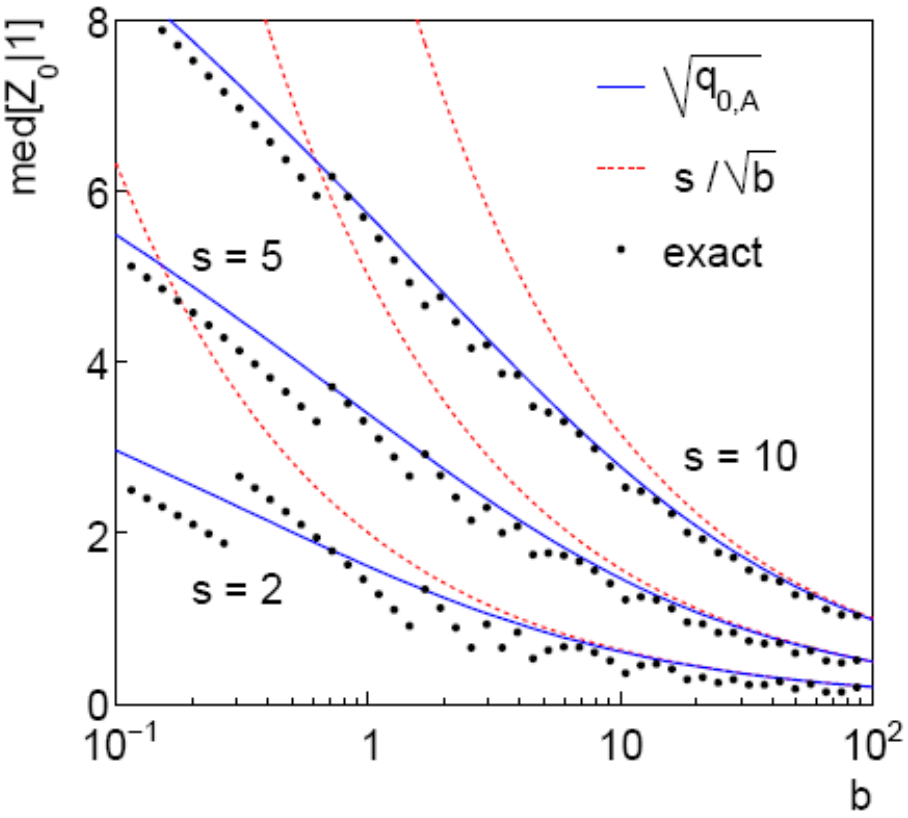
To find median$[Z_0|s+b]$, let $n \to s + b$ (i.e., the Asimov data set):

$$\text{median}[Z_0|s+b] \approx \sqrt{2 \left((s+b) \ln(1 + s/b) - s \right)}$$

This reduces to $s/\sqrt{b}$ for s << b.

$n \sim \text{Poisson}(\mu s + b)$, median significance, assuming $\mu = 1$, of the hypothesis $\mu = 0$

CCGV, arXiv:1007.1727



"Exact" values from MC, jumps due to discrete data.

Asimov $\sqrt{q_{0,A}}$ good approx. for broad range of s, b.

$s/\sqrt{b}$ only good for $s \ll b$.

Using likelihood ratio L_{s+b}/L_b

Many searches at the Tevatron have used the statistic

$$q = -2 \ln \frac{L_{s+b}}{L_b}$$

likelihood of $\mu = 1$ model (s+b)

likelihood of $\mu = 0$ model (bkg only)

This can be written

$$q = -2 \ln \frac{L(\mu = 1, \hat{\hat{\boldsymbol{\theta}}}(1))}{L(\mu = 0, \hat{\hat{\boldsymbol{\theta}}}(0))} = -2 \ln \lambda(1) + 2 \ln \lambda(0)$$

Wald approximation for L_{s+b}/L_b

Assuming the Wald approximation, q can be written as

$$q = \frac{(\hat{\mu} - 1)^2}{\sigma^2} - \frac{\hat{\mu}^2}{\sigma^2} = \frac{1 - 2\hat{\mu}}{\sigma^2}$$

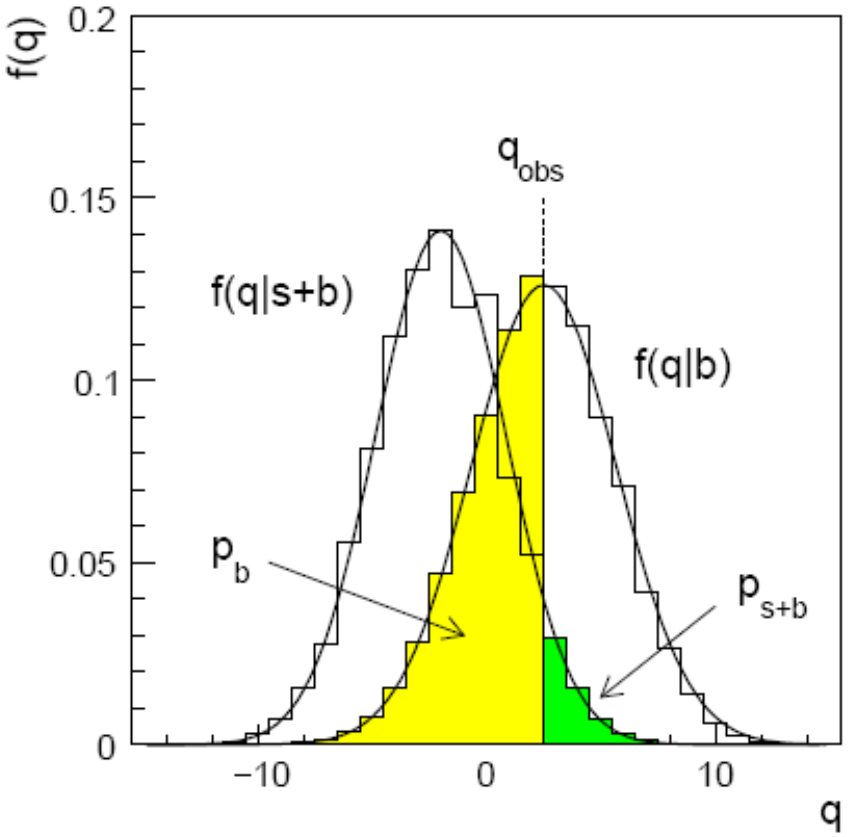
i.e. q is Gaussian distributed with mean and variance of

$$E[q] = \frac{1 - 2\mu}{\sigma^2} \qquad V[q] = \frac{4}{\sigma^2}$$

To get σ^2 use 2nd derivatives of ln$L$ with Asimov data set.

Example with L_{s+b}/L_b

Consider again $n \sim \text{Poisson}(\mu s + b)$, $m \sim \text{Poisson}(\tau b)$
$b = 20$, $s = 10$, $\tau = 1$.



So even for smallish data sample, Wald approximation can be useful; no MC needed.

Summary on asymptotic formulae

Asymptotic distributions of profile LR applied to an LHC search.

Wilks: $f(q_\mu|\mu)$ for p-value of μ.

Wald approximation for $f(q_\mu|\mu')$.

"Asimov" data set used to estimate median q_μ for sensitivity.

Gives σ of distribution of estimator of μ.

Asymptotic formulae especially useful for estimating sensitivity in high-dimensional parameter space.

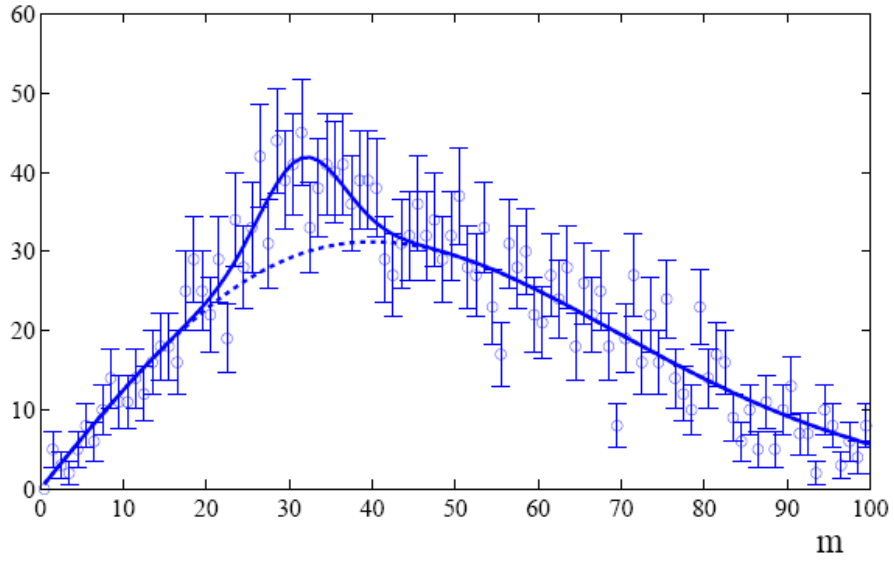
Can always check with MC for very low data samples and/or when precision crucial.

The Look-Elsewhere Effect

Eilam Gross and Ofer Vitells, arXiv:1005.1891 (→ EPJC)

Suppose a model for a mass distribution allows for a peak at a mass m with amplitude μ.

The data show a bump at a mass m_0.



How consistent is this with the no-bump ($\mu = 0$) hypothesis?

Eilam Gross and Ofer Vitells, arXiv:1005.1891 (→EPJC)

p-value for fixed mass

First, suppose the mass m_0 of the peak was specified a priori.

Test consistency of bump with the no-signal ($\mu = 0$) hypothesis with e.g. likelihood ratio

$$t_{\text{fix}} = -2 \ln \frac{L(0, m_0)}{L(\hat{\mu}, m_0)}$$

where "fix" indicates that the mass of the peak is fixed to m_0.

The resulting p-value

$$p_{\text{fix}} = \int_{t_{\text{fix,obs}}}^{\infty} f(t_{\text{fix}}|0) dt_{\text{fix}}$$

gives the probability to find a value of t_{fix} at least as great as observed at the specific mass m_0.

p-value for floating mass

But suppose we did not know where in the distribution to expect a peak.

What we want is the probability to find a peak at least as significant as the one observed anywhere in the distribution.

Include the mass as an adjustable parameter in the fit, test significance of peak using

$$t_{\text{float}} = -2 \ln \frac{L(0)}{L(\hat{\mu}, \hat{m})}$$

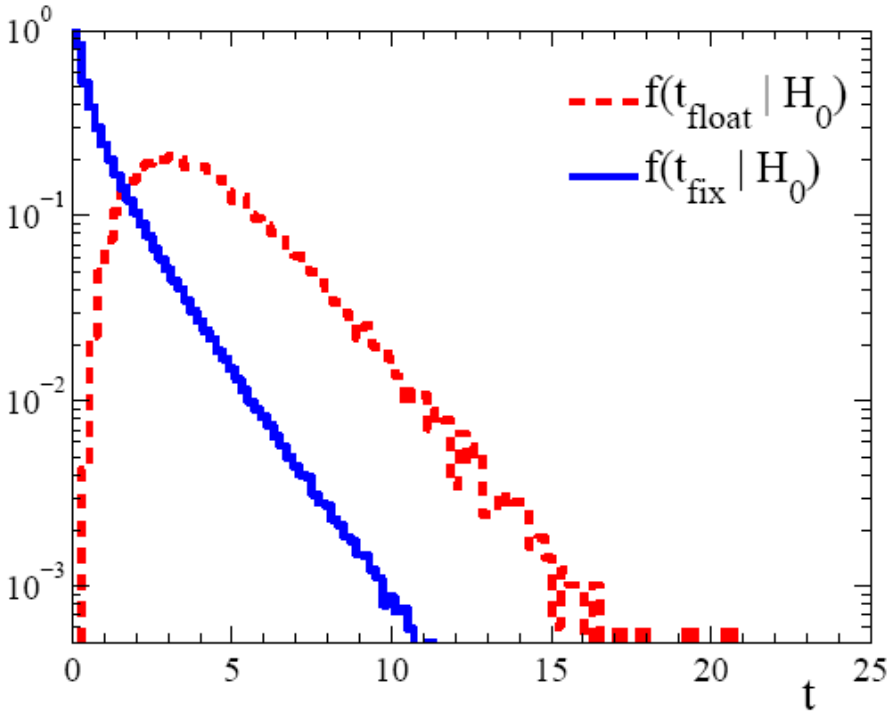
(Note m does not appear in the $\mu = 0$ model.)

$$p_{\text{float}} = \int_{t_{\text{float,obs}}}^{\infty} f(t_{\text{float}}|0) dt_{\text{float}}$$

Distributions of t_{fix}, t_{float}

For a sufficiently large data sample, t_{fix} ~chi-square for 1 degree of freedom (Wilks' theorem).

For t_{float} there are two adjustable parameters, μ and m, and naively Wilks theorem says t_{float} ~ chi-square for 2 d.o.f.



In fact Wilks' theorem does not hold in the floating mass case because on of the parameters (m) is not-defined in the $\mu = 0$ model.

So getting t_{float} distribution is more difficult.

Eilam Gross and Ofer Vitells, arXiv:1005.1891 (→EPJC)

Trials factor

We would like to be able to relate the *p*-values for the fixed and floating mass analyses (at least approximately).

Gross and Vitells (arXiv:1005.1891) show that the "trials factor" can be approximated by

$$F_{\text{trials}} \equiv \frac{p_{\text{float}}}{p_{\text{fix}}} \approx 1 + \sqrt{\frac{\pi}{2}} \langle \mathcal{N} \rangle Z_{\text{fix}}$$

where ‹*N*› = average number of "upcrossings" of −2ln*L* in fit range and

$$Z_{\text{fix}} = \Phi^{-1}(1 - p_{\text{fix}}) = \sqrt{t_{\text{fix}}}$$

is the significance for the fixed mass case.

So we can either carry out the full floating-mass analysis (e.g. use MC to get *p*-value), or do fixed mass analysis and apply a correction factor (much faster than MC).

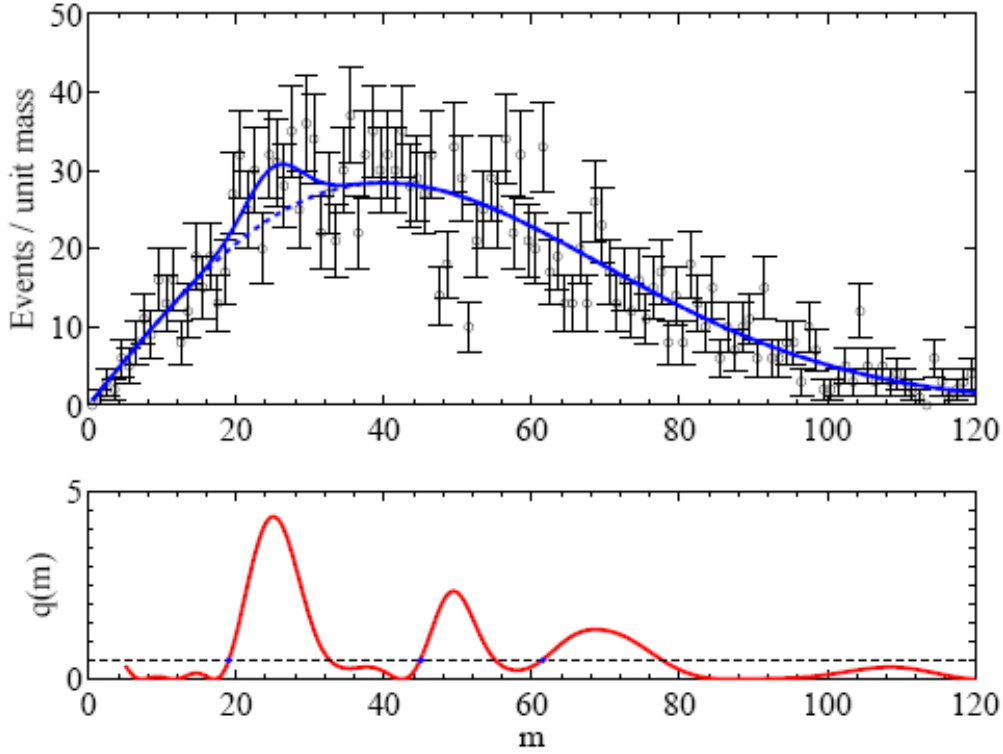
Eilam Gross and Ofer Vitells, arXiv:1005.1891 (→EPJC)

Upcrossings of $-2\ln L$

The Gross-Vitells formula for the trials factor requires the mean number "upcrossings" of $-2\ln L$ in the fit range based on fixed threshold.

$$P(q_0 > u)$$
$$\leq E[N_u] + P(q_0(0) > u)$$
$$= \mathcal{N}_1 e^{-u/2} + \frac{1}{2} P(\chi_1^2 > u)$$

estimate with MC at low reference level

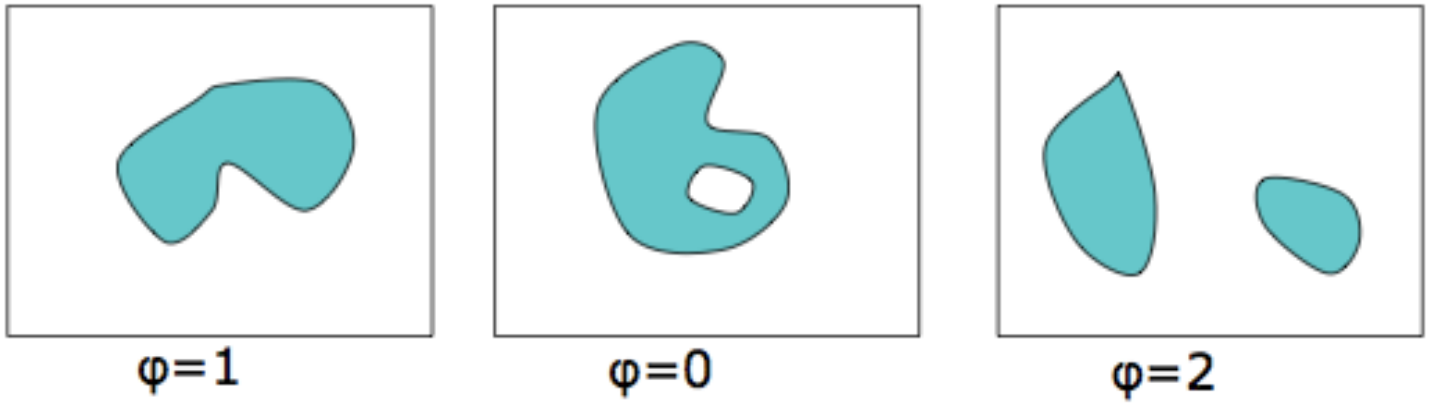


Multidimensional look-elsewhere effect

Generalization to multiple dimensions: number of upcrossings replaced by expectation of Euler characteristic:

$$\mathrm{E}[\varphi(A_u)] = \sum_{d=0}^{n} \mathcal{N}_d \rho_d(u)$$

- Number of disconnected components minus number of `holes'



Applications: astrophysics (coordinates on sky), search for resonance of unknown mass and width, ...

Combination of channels

For a set of independent decay channels, full likelihood function is product of the individual ones:

$$L(\mu, \boldsymbol{\theta}) = \prod_i L_i(\mu, \boldsymbol{\theta}_i)$$

For combination need to form the full function and maximize to find estimators of μ, $\boldsymbol{\theta}$.

→ ongoing ATLAS/CMS effort with RooStats framework

https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome

Trick for median significance: estimator for μ is equal to the Asimov value μ' for all channels separately, so for combination,

$$\lambda_{\rm A}(\mu) = \prod_i \lambda_{{\rm A},i}(\mu) \qquad \text{where} \qquad \lambda_{{\rm A},i}(\mu) = \frac{L_i(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L_i(\mu', \boldsymbol{\theta})}$$

RooStats

G. Schott
PHYSTAT2011

a collaborative project with contributors from ATLAS, CMS and ROOT aimed to provide & consolidate statistical tools needed by LHC

- using same tools: compare easily results across experiments
  - not only desirable but necessary for combinations

RooStats is built on top of the RooFit toolkit :

- data modelling language (for PDFs, likelihoods, ...)

RooFit Workspaces

G. Schott
PHYSTAT2011

RooWorkspace class of RooFit: possibility to save it to a ROOT file

- very good for electronic publication of data and likelihood function
- and greatly help for combination (that's the format agreed to share between Atlas & CMS)

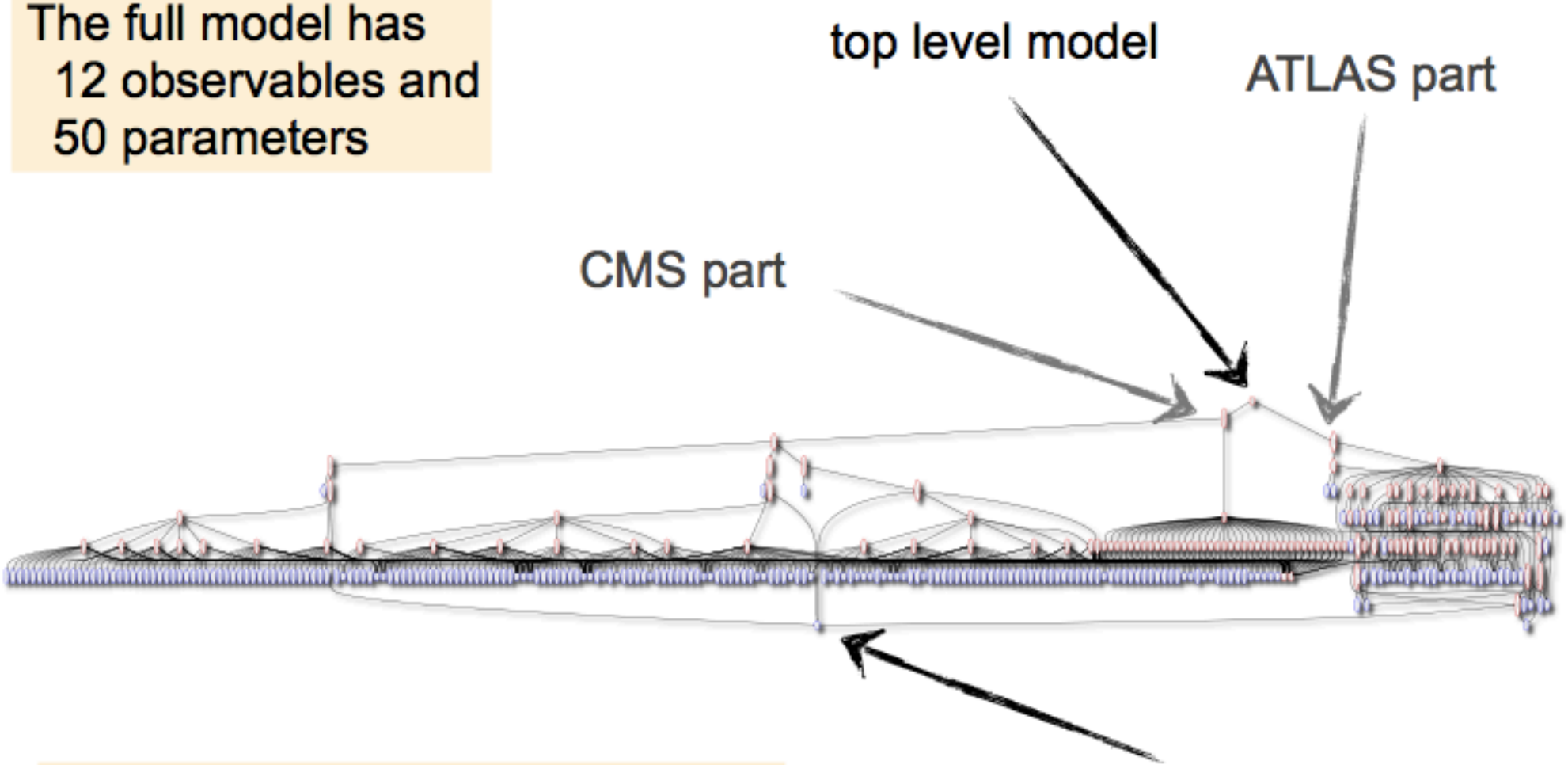
```
RooWorkspace w("w","joint workspace") ;
// Import top-level pdfs and all their components, variables
w.import("channelA.root:w:pdfA",RenameAllVariablesExcept("A","mhiggs"))
w.import("channelB.root:w:pdfB",RenameVariable("mH","mhiggs")) ;
w.import("channelC.root:w:pdfC") ;
// Construct joint pdf
w.factory("SIMUL::joint(chan[A,B,C],A=pdfA,B=pdfB,C=pdfC)") ;
```

Able to construct full likelihood for combination of channels (or experiments).

K. Cranmer
PHYSTAT2011

Combined ATLAS/CMS Higgs search

The full model has
12 observables and
50 parameters



At this point, no correlated systematics across experiments

$$\mu = \frac{\sigma BR}{\sigma_{SM} BR_{SM}}$$

Summary of the rest

Progress on related issues for LHC discovery:

Look elsewhere effect (Gross and Vitells)

New software for combinations (and other things): RooStats

Needed:

More work on how to parametrize models so as to include a level of flexibility commensurate with the real systematic uncertainty, together with ideas on how to constrain this flexibility experimentally (control measurements).

Extra slides

Negatively Biased Relevant Subsets

Consider again $x \sim \text{Gauss}(\mu, \sigma)$ and use this to find limit for μ.

We can find the conditional probability for the limit to cover μ given x in some restricted range, e.g., $x < c$ for some constant c.

This conditional coverage probability may be greater or less than $1 - \alpha$ for different values of μ (the value of which is unkown).

But suppose that the conditional coverage is less than $1 - \alpha$ for *all* values of μ. The region of x where this is true is a *Negatively Biased Relevant Subset*.

Recent studies by Bob Cousins (CMS) and
Ofer Vitells (ATLAS) related to earlier publications,
especially, R. Buehler, Ann. Math. Sci., 30 (4) (1959) 845.

Betting Games

So what's wrong if the limit procedure has NBRS?

Suppose you observe x, construct the confidence interval and assert that an interval thus constructed covers the true value of the parameter with probability $1 - \alpha$.

This means you should be willing to accept a bet at odds $\alpha : 1 - \alpha$ that the interval covers the true parameter value.

Suppose your opponent accepts the bet if x is in the NBRS, and declines the bet otherwise. On average, you lose, regardless of the true (and unknown) value of μ.

With the "naive" unconstrained limit, if your opponent only accepts the bet when $x < -1.64\sigma$, (all values of μ excluded) you always lose!

(Recall the unconstrained limit based on the likelihood ratio never excludes $\mu = 0$, so if that value is true, you do not lose.)

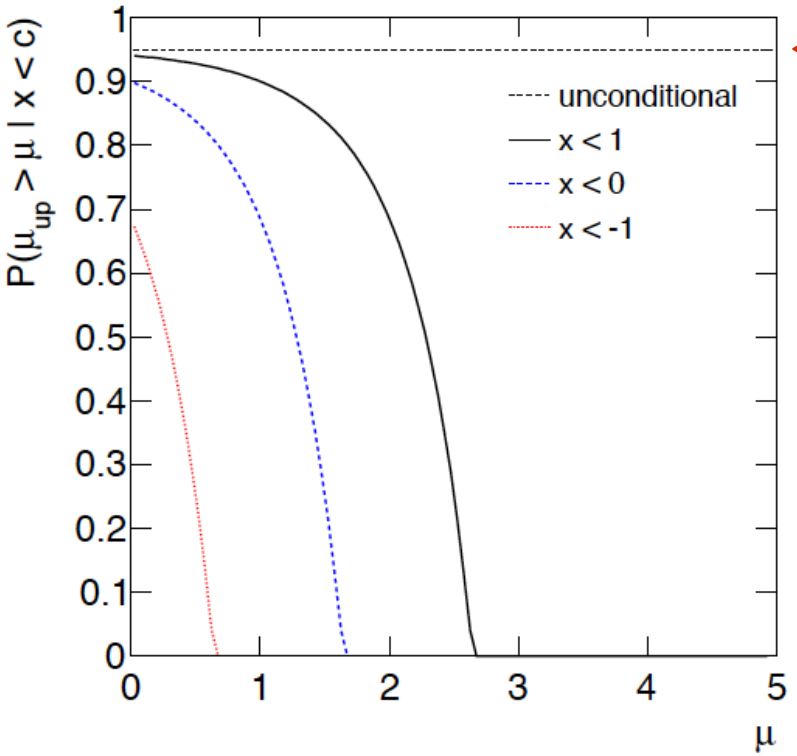
NBRS for unconstrained upper limit

For the unconstrained upper limit (i.e., CL_{s+b}) the conditional probability for the limit to cover μ given $x < c$ is:

$$P(\mu_{\rm up} > \mu | x < c) = \frac{1 - \alpha - \Phi\left(\frac{\mu - c}{\sigma}\right)}{1 - \Phi\left(\frac{\mu - c}{\sigma}\right)}$$

Maximum wrt μ is less than $1-\alpha$ → Negatively biased relevant subsets.

N.B. $\mu = 0$ is never excluded for unconstrained limit based on likelihood-ratio test, so at that point coverage = 100%, hence no NBRS.



← $1 - \alpha$

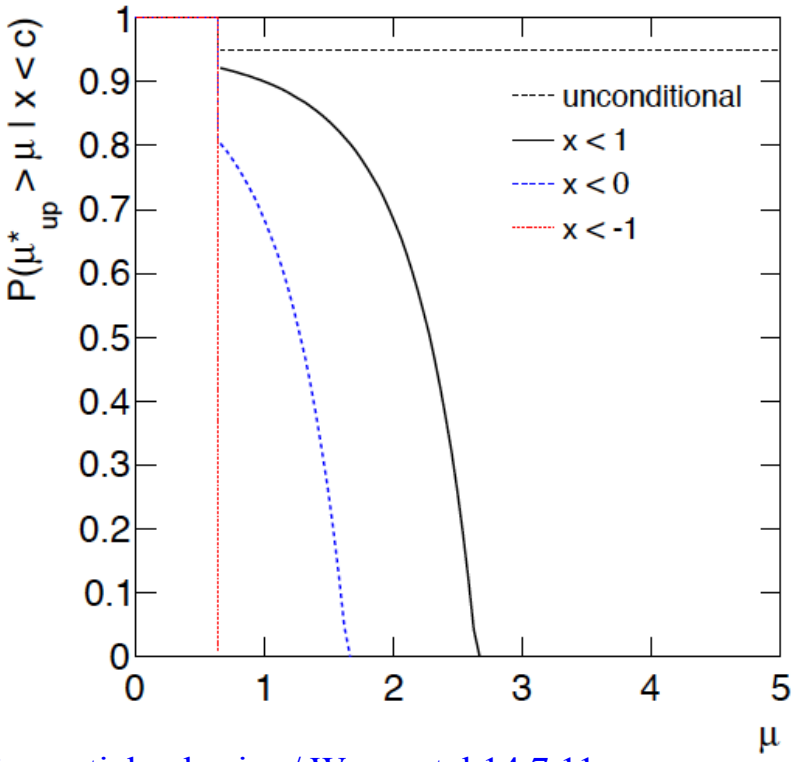
(Adapted) NBRS for PCL

For PCL, the conditional probability to cover μ given $x < c$ is:

$$P(\mu^*_{\rm up} > \mu | x < c) = \begin{cases} 1 & \mu < \mu_{\rm min}, \\ \frac{1-\alpha-\Phi\left(\frac{\mu-c}{\sigma}\right)}{1-\Phi\left(\frac{\mu-c}{\sigma}\right)} & \text{otherwise.} \end{cases}$$

Coverage goes to 100% for $\mu < \mu_{\rm min}$, therefore no NBRS.

Note one does not have max conditional coverage $\geq 1-\alpha$ for all $\mu > \mu_{\rm min}$ ("adapted conditional coverage"). But if one conditions on μ, no limit would satisfy this.



← $1-\alpha$

Conditional coverage for CLs, F-C