Statistical Methods for Particle Physics

Lecture 3: asymptotics I; Asimov data set

https://indico.weizmann.ac.il//conferenceDisplay.py?confId=52

Statistical Inference for Astro and Particle Physics Workshop
Weizmann Institute, Rehovot
March 8-12, 2015

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

Outline for Monday – Thursday

(GC = Glen Cowan, KC = Kyle Cranmer)

Monday 9 March

GC: probability, random variables and related quantities

KC: parameter estimation, bias, variance, max likelihood

Tuesday 10 March

KC: building statistical models, nuisance parameters

GC: hypothesis tests I, *p*-values, multivariate methods

Wednesday 11 March

KC: hypothesis tests 2, composite hyp., Wilks', Wald's thm.

GC: asympotics 1, Asimov data set, sensitivity

Thursday 12 March:

KC: confidence intervals, asymptotics 2

GC: unfolding

Recap of frequentist 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.)

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

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, EPJC 71 (2011) 1554; arXiv:1007.1727.

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 $1 - \alpha$.

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

Put critical region where data are likely to be under assumption of the relevant alternative to the μ that's being tested.

Test $\mu = 0$, alternative is $\mu > 0$: test for discovery.

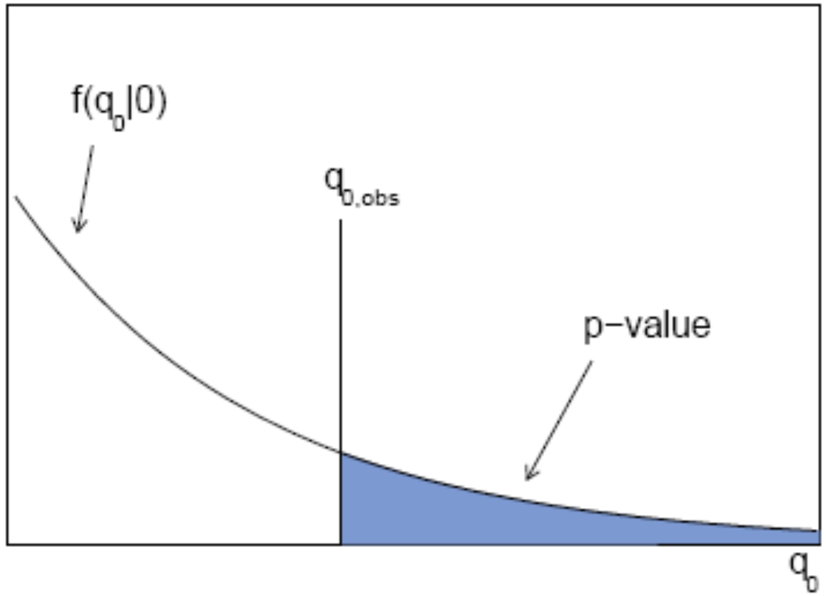
Test $\mu = \mu_0$, alternative is $\mu = 0$: testing all μ_0 gives upper limit.

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,

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

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

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

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

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

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.

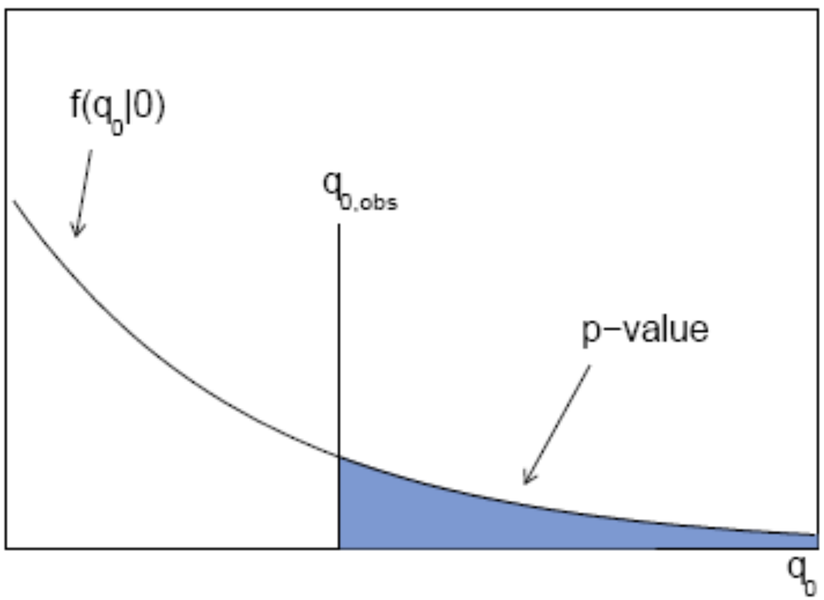
Note that even though here physically $\mu \geq 0$, we allow $\hat{\mu}$ to be negative. In large sample limit its distribution becomes Gaussian, and this will allow us to write down simple expressions for distributions of our test statistics.

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

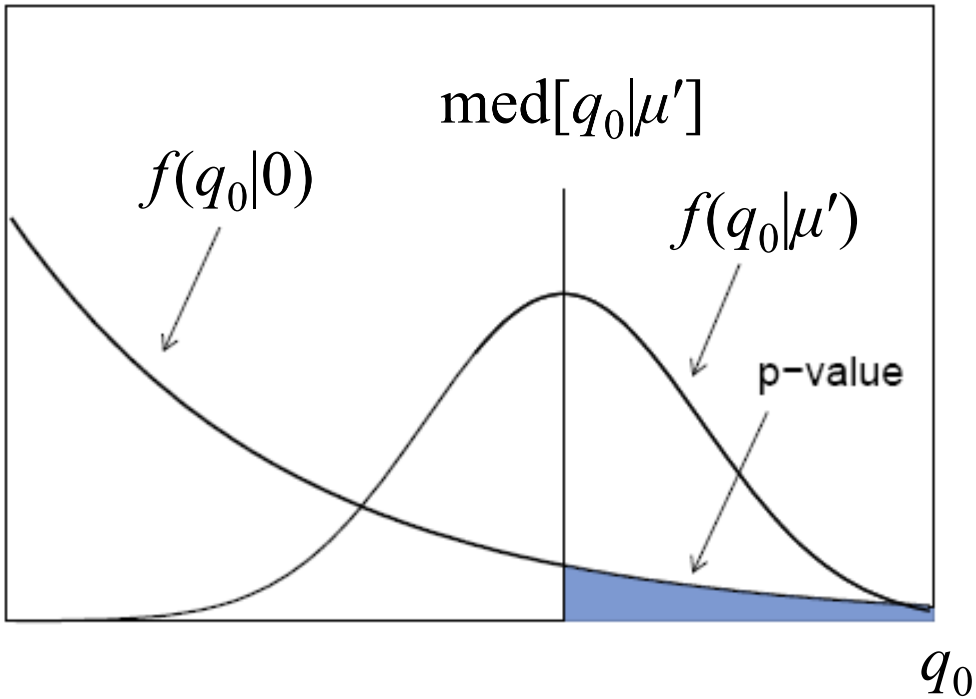


From p-value get equivalent significance,

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

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 $f(q_0|\mu')$,

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)

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

sample size

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

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

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):

$$
\begin{aligned}
n_i &= \mu' s_i + b_i \\
m_i &= u_i
\end{aligned}
$$

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

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

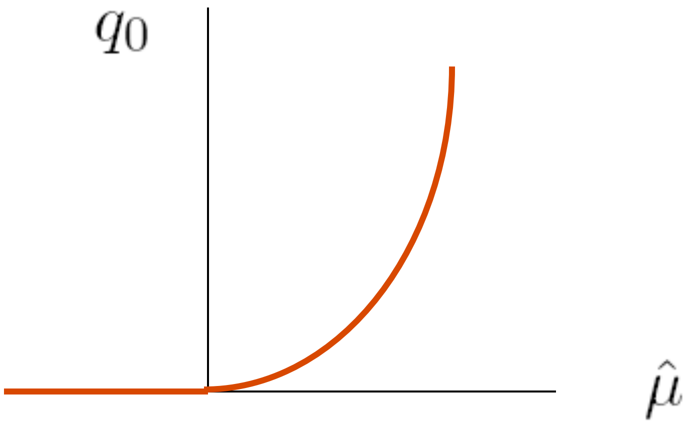
$$-2\ln\lambda_{\mathrm{A}}(\mu) = \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda$$

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.,

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

$$\mathrm{med}[Z_0] = \sqrt{-2\ln\lambda_{\mathrm{A}}(0)}$$

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

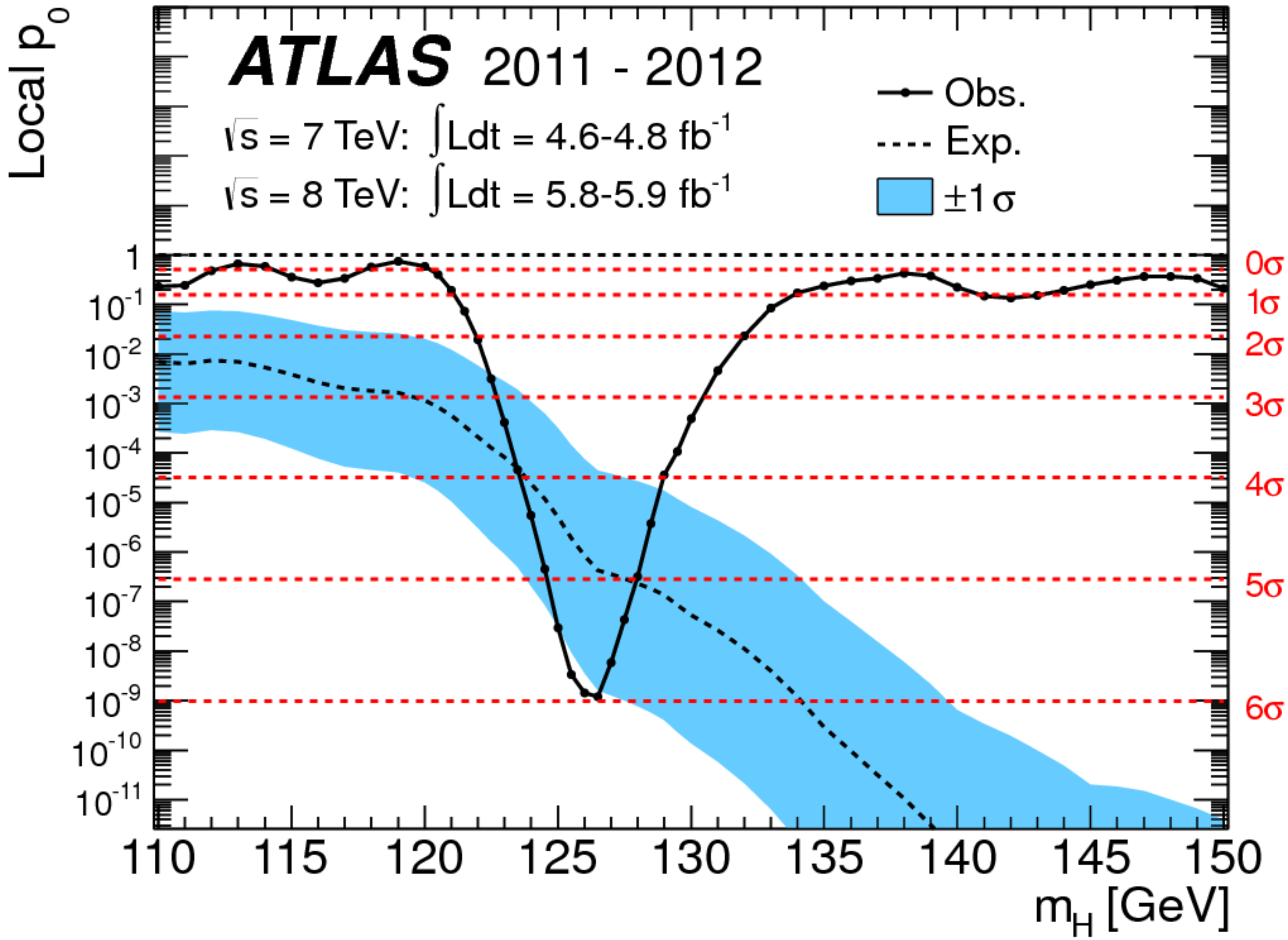
$$p_0 = 1 - F(q_0|0)$$

Therefore the discovery significance Z is simply

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

Example of a *p*-value

ATLAS, Phys. Lett. B 716 (2012) 1-29



Profile likelihood ratio 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.

Note also here we allow the estimator for μ be negative (but $\hat{\mu}s_i + b_i$ must be positive).

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

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

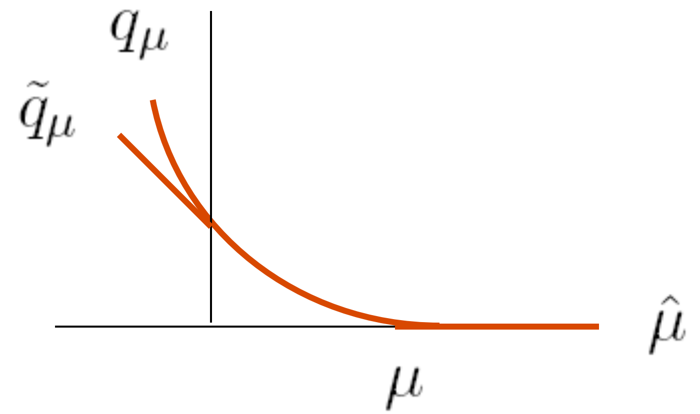
Performance not identical to but very close to q_μ (of previous slide).

q_μ is simpler in important ways.

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 $\tilde{q}_\mu$

$$f(\tilde{q}_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(\tilde{q}_\mu)$$

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

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

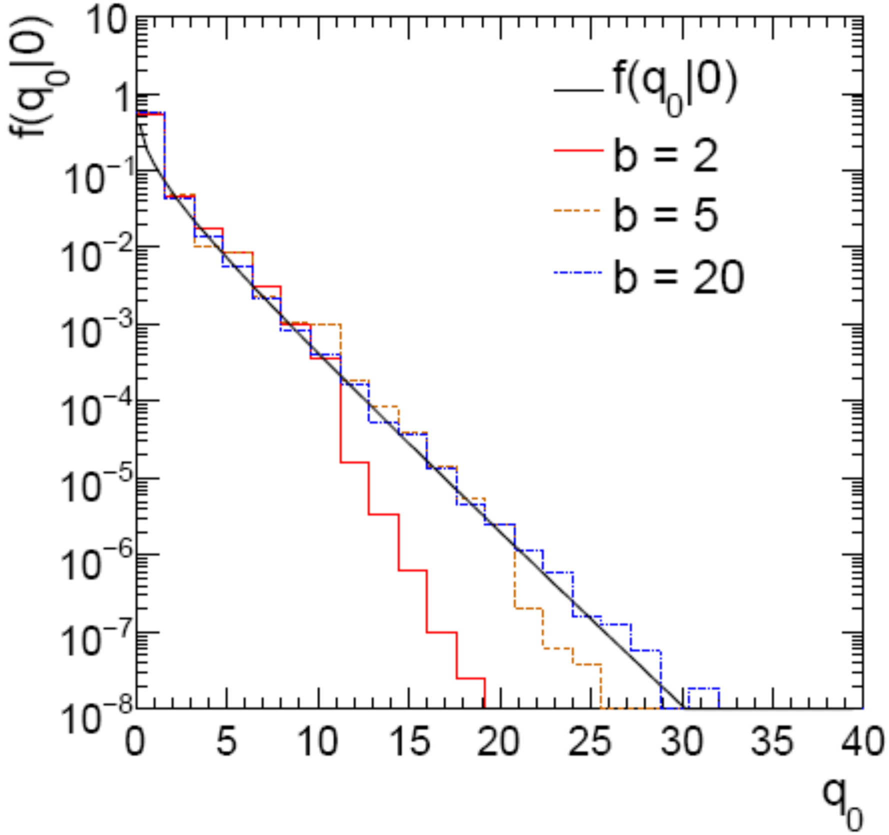
Monte Carlo test of asymptotic formula

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

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

Here take $\tau = 1$.

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

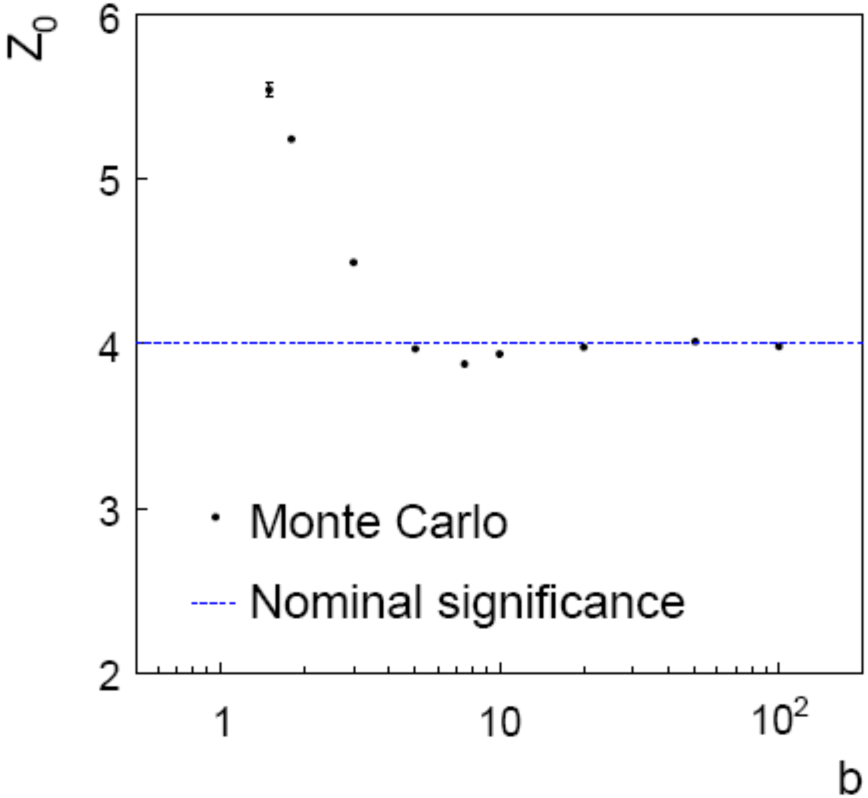


Monte Carlo test of asymptotic formulae

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.

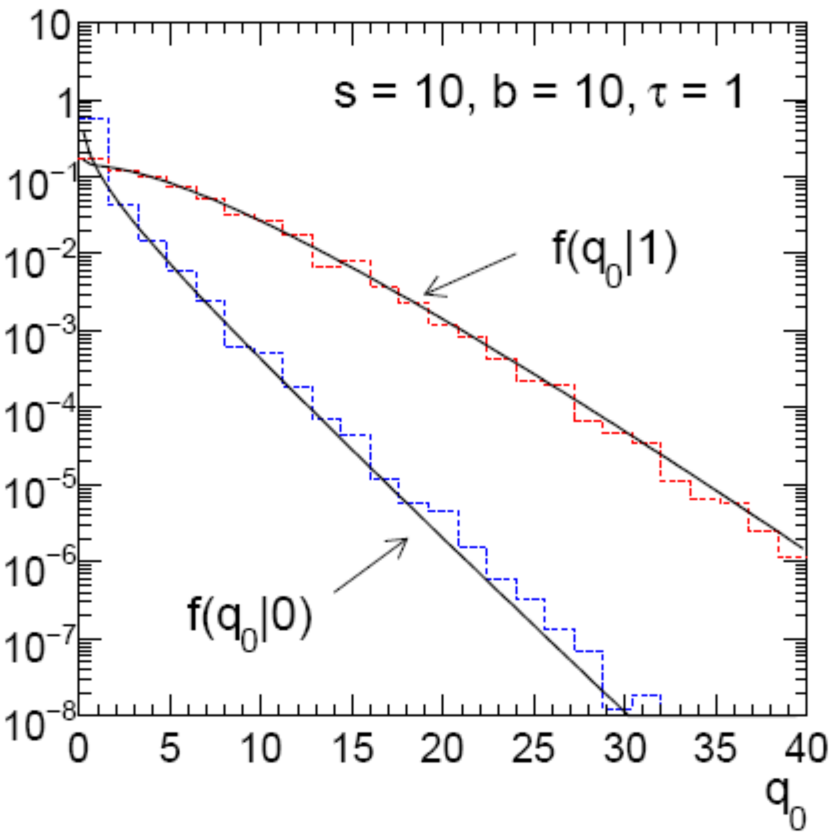
Then slight overshoot before rapidly converging to MC value.

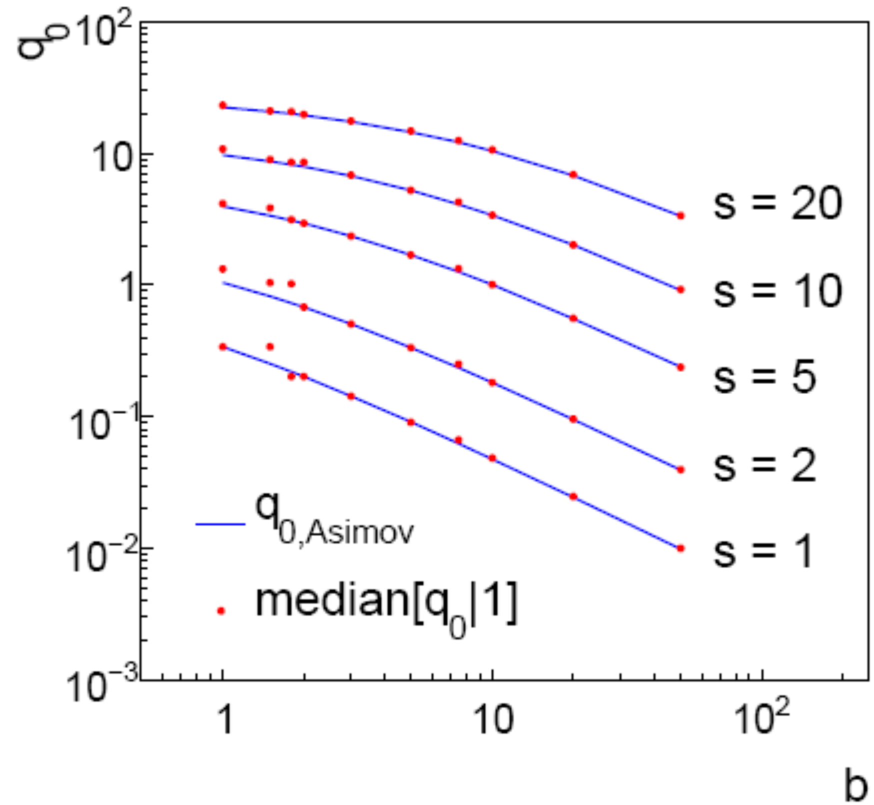


Monte Carlo test of asymptotic formulae

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

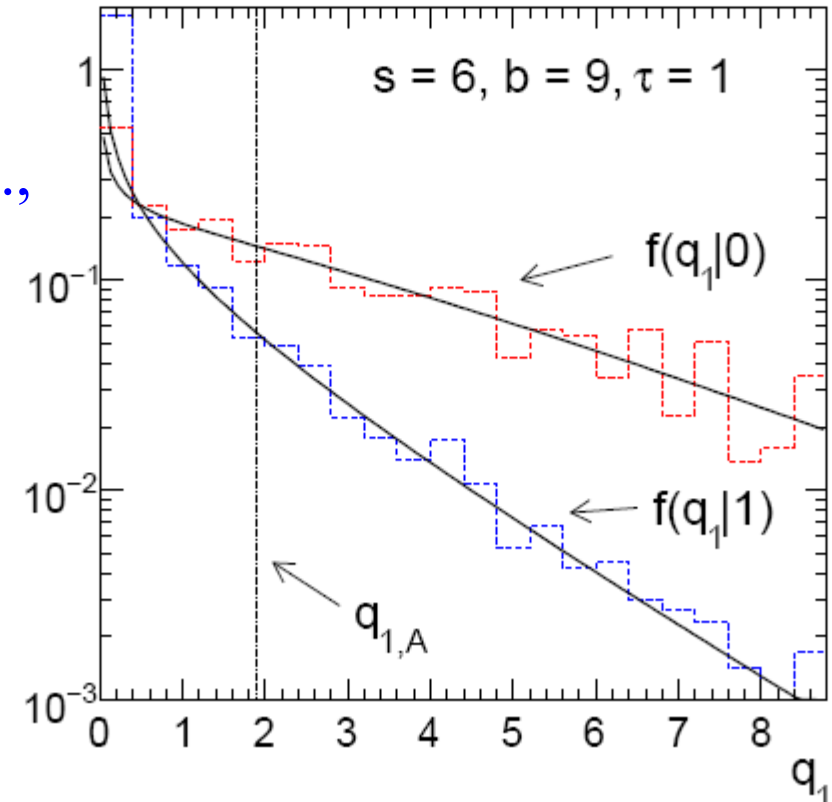
Consider again $n \sim$ Poisson $(\mu s + b)$, $m \sim$ Poisson(τb)
Use q_μ to find p-value of hypothesized μ values.

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".

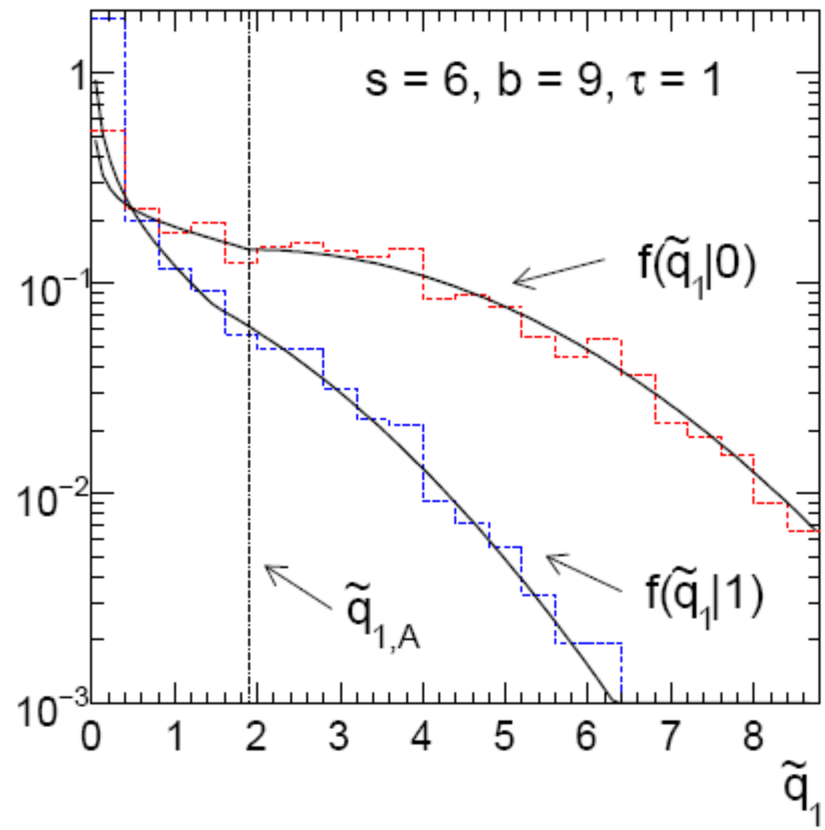
Here asymptotic formulae good for $s = 6$, $b = 9$.



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.



Expected discovery significance for counting experiment with background uncertainty

I. Discovery sensitivity for counting experiment with b known:

(a) $$\frac{s}{\sqrt{b}}$$

(b) Profile likelihood ratio test & Asimov: $$\sqrt{2\left((s+b)\ln\left(1+\frac{s}{b}\right)-s\right)}$$

II. Discovery sensitivity with uncertainty in b, σ_b:

(a) $$\frac{s}{\sqrt{b+\sigma_b^2}}$$

(b) Profile likelihood ratio test & Asimov:

$$\left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right]-\frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$

Counting experiment with known background

Count a number of events $n \sim$ Poisson($s+b$), where

s = expected number of events from signal,

b = expected number of background events.

To test for discovery of signal compute p-value of $s = 0$ hypothesis,

$$p = P(n \geq n_{\rm obs}|b) = \sum_{n=n_{\rm obs}}^{\infty} \frac{b^n}{n!} e^{-b} = 1 - F_{\chi^2}(2b; 2n_{\rm obs})$$

Usually convert to equivalent significance: $Z = \Phi^{-1}(1-p)$
where Φ is the standard Gaussian cumulative distribution, e.g.,
$Z > 5$ (a 5 sigma effect) means $p < 2.9 \times 10^{-7}$.

To characterize sensitivity to discovery, give expected (mean or median) Z under assumption of a given s.

$s/\sqrt{b}$ for expected discovery significance

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

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

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

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

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

Better approximation for significance

Poisson likelihood for parameter s is

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

For now no nuisance params.

To test for discovery use profile likelihood ratio:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{s} \geq 0\,, \\ 0 & \hat{s} < 0\,. \end{cases} \qquad \lambda(s) = \frac{L(s, \hat{\hat{\theta}}(s))}{L(\hat{s}, \hat{\theta})}$$

So the likelihood ratio statistic for testing $s = 0$ is

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

Approximate Poisson significance (continued)

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

$$Z = \sqrt{2\left(n \ln \frac{n}{b} + b - n\right)} \quad \text{for } n > b \text{ and } Z = 0 \text{ otherwise.}$$

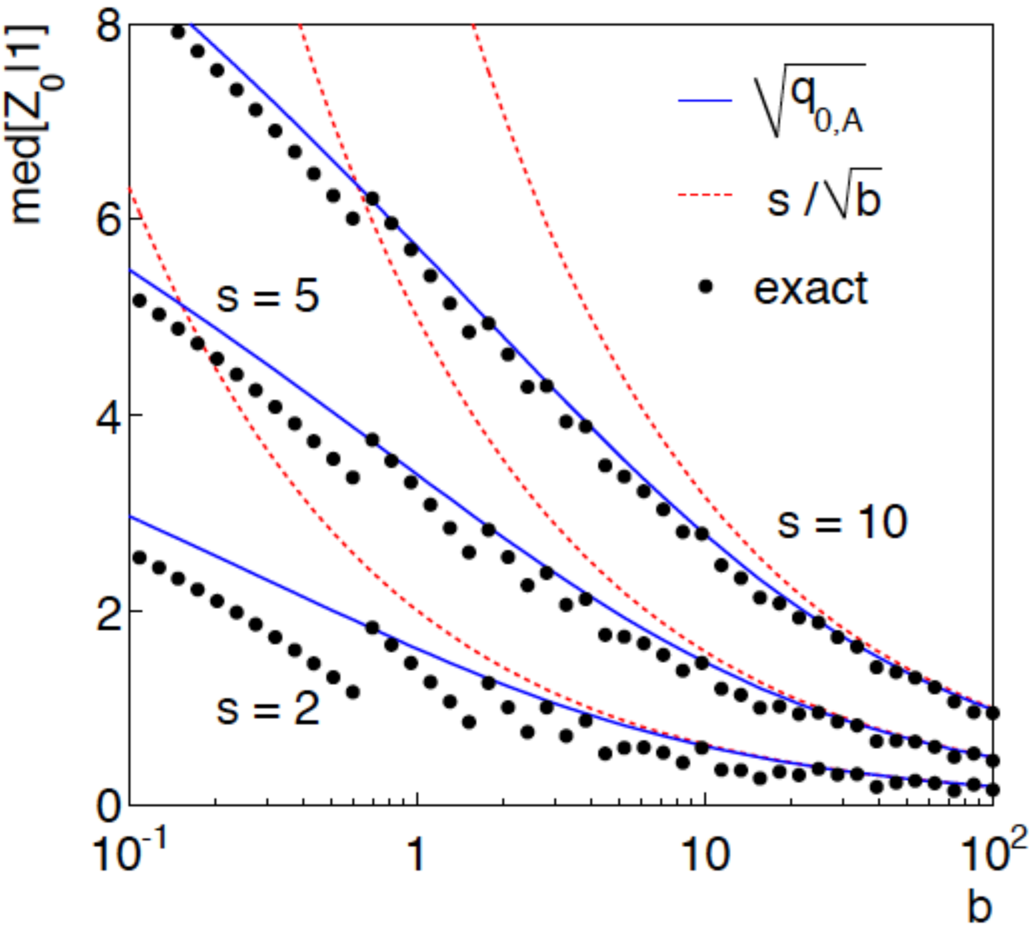
To find median[$Z|s$], let $n \to s + b$ (i.e., the Asimov data set):

$$Z_{\mathrm{A}} = \sqrt{2\left((s+b)\ln\left(1 + \frac{s}{b}\right) - s\right)}$$

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

$n \sim$ Poisson($s+b$), median significance, assuming s, of the hypothesis $s = 0$

CCGV, EPJC 71 (2011) 1554, 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 $s \ll b$.

Extending $s/\sqrt{b}$ to case where b uncertain

The intuitive explanation of $s/\sqrt{b}$ is that it compares the signal, s, to the standard deviation of n assuming no signal, $\sqrt{b}$.

Now suppose the value of b is uncertain, characterized by a standard deviation σ_b.

A reasonable guess is to replace $\sqrt{b}$ by the quadratic sum of $\sqrt{b}$ and σ_b, i.e.,

$$\mathrm{med}[Z|s] = \frac{s}{\sqrt{b + \sigma_b^2}}$$

This has been used to optimize some analyses e.g. where σ_b cannot be neglected.

Profile likelihood with *b* uncertain

This is the well studied "on/off" problem: Cranmer 2005; Cousins, Linnemann, and Tucker 2008; Li and Ma 1983,...

Measure two Poisson distributed values:

$n \sim \text{Poisson}(s+b)$ (primary or "search" measurement)

$m \sim \text{Poisson}(\tau b)$ (control measurement, τ known)

The likelihood function is

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

Use this to construct profile likelihood ratio (b is nuisance parmeter):

$$\lambda(0) = \frac{L(0, \hat{\hat{b}}(0))}{L(\hat{s}, \hat{b})}$$

Ingredients for profile likelihood ratio

To construct profile likelihood ratio from this need estimators:

$$\hat{s} = n - m/\tau \,,$$

$$\hat{b} = m/\tau \,,$$

$$\hat{\hat{b}}(s) = \frac{n + m - (1+\tau)s + \sqrt{(n + m - (1+\tau)s)^2 + 4(1+\tau)sm}}{2(1+\tau)} \,.$$

and in particular to test for discovery ($s = 0$),

$$\hat{\hat{b}}(0) = \frac{n+m}{1+\tau}$$

Asymptotic significance

Use profile likelihood ratio for q_0, and then from this get discovery significance using asymptotic approximation (Wilks' theorem):

$$Z = \sqrt{q_0}$$

$$= \left[-2 \left(n \ln \left[\frac{n+m}{(1+\tau)n} \right] + m \ln \left[\frac{\tau(n+m)}{(1+\tau)m} \right] \right) \right]^{1/2}$$

for $n > \hat{b}$ and $Z = 0$ otherwise.

Essentially same as in:

Robert D. Cousins, James T. Linnemann and Jordan Tucker, NIM A 595 (2008) 480–501; arXiv:physics/0702156.

Tipei Li and Yuqian Ma, Astrophysical Journal 272 (1983) 317–324.

Asimov approximation for median significance

To get median discovery significance, replace n, m by their expectation values assuming background-plus-signal model:

$$n \to s + b$$

$$m \to \tau b$$

$$Z_{\rm A} = \left[-2 \left((s+b) \ln \left[\frac{s + (1+\tau) b}{(1+\tau)(s+b)} \right] + \tau b \ln \left[1 + \frac{s}{(1+\tau) b} \right] \right) \right]^{1/2}$$

Or use the variance of $\hat{b} = m/\tau$, $V[\hat{b}] \equiv \sigma_b^2 = \dfrac{b}{\tau}$, to eliminate τ:

$$\boxed{Z_{\rm A} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}}$$

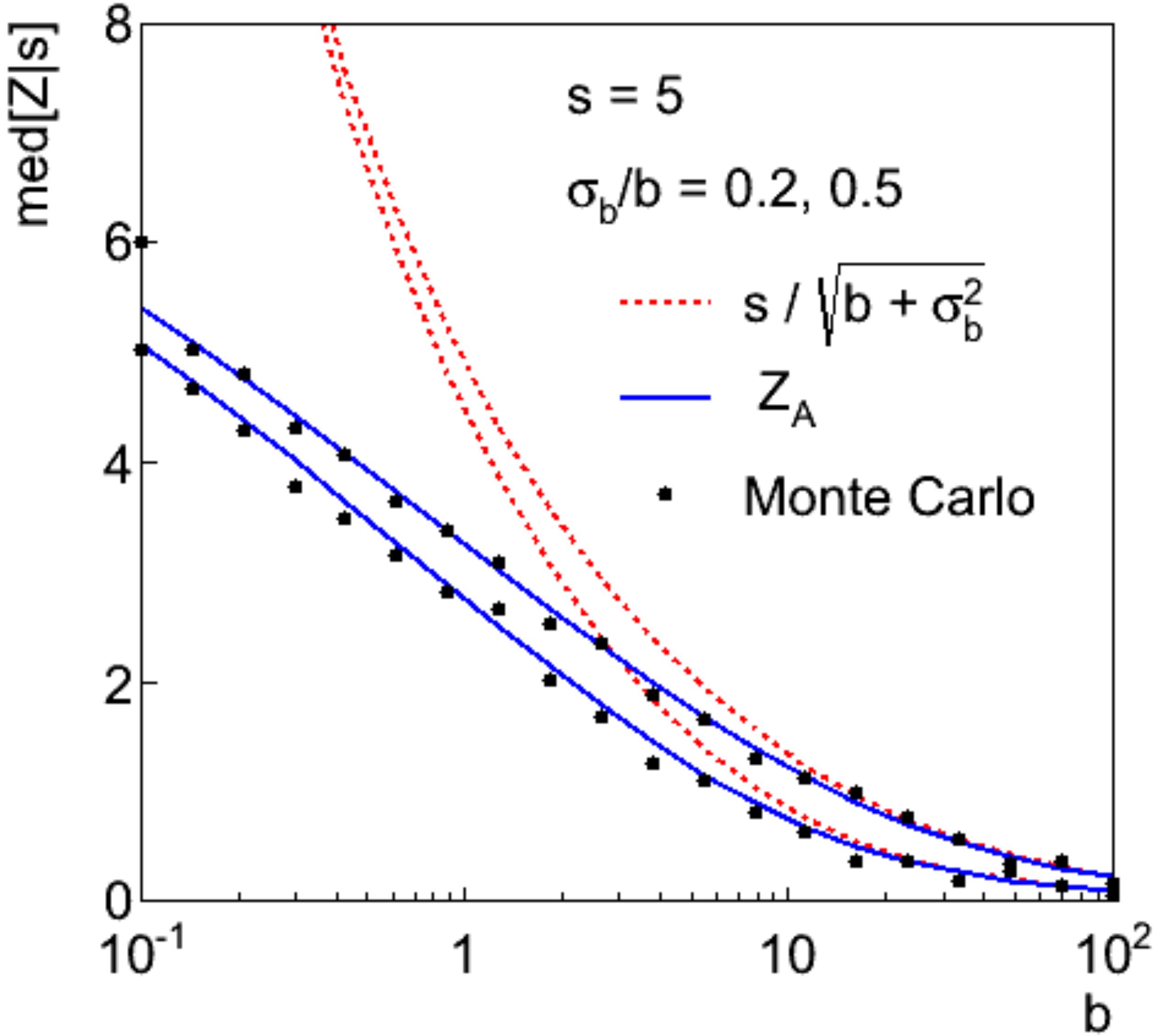
Limiting cases

Expanding the Asimov formula in powers of s/b and σ_b^2/b (= $1/\tau$) gives

$$Z_{\mathrm{A}} = \frac{s}{\sqrt{b + \sigma_b^2}} \left(1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)$$

So the "intuitive" formula can be justified as a limiting case of the significance from the profile likelihood ratio test evaluated with the Asimov data set.

Testing the formulae: $s = 5$



Using sensitivity to optimize a cut

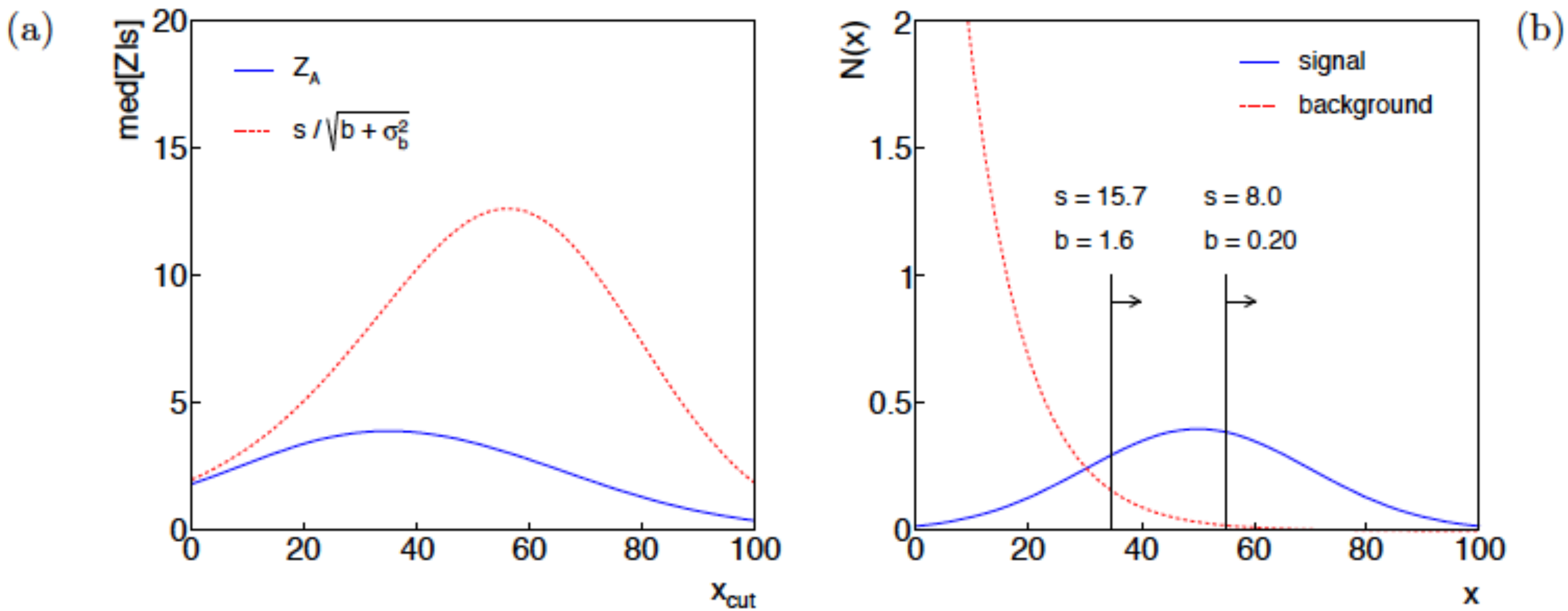


Figure 1: (a) The expected significance as a function of the cut value $x_{\rm cut}$; (b) the distributions of signal and background with the optimal cut value indicated.

Summary on discovery sensitivity

Simple formula for expected discovery significance based on profile likelihood ratio test and Asimov approximation:

$$Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

For large b, all formulae OK.

For small b, $s/\sqrt{b}$ and $s/\sqrt{(b+\sigma_b^2)}$ overestimate the significance.

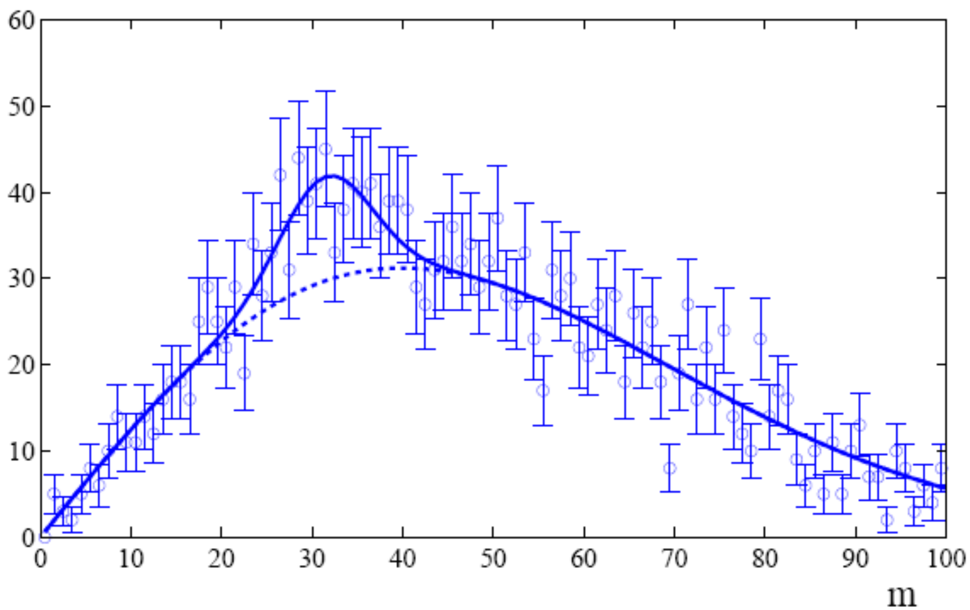
Could be important in optimization of searches with low background.

Formula maybe also OK if model is not simple on/off experiment, e.g., several background control measurements (checking this).

The Look-Elsewhere Effect

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?

Local p-value

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

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

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

The resulting p-value

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

gives the probability to find a value of $t_{\rm fix}$ at least as great as observed at the specific mass m_0 and is called the local p-value.

Global p-value

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

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

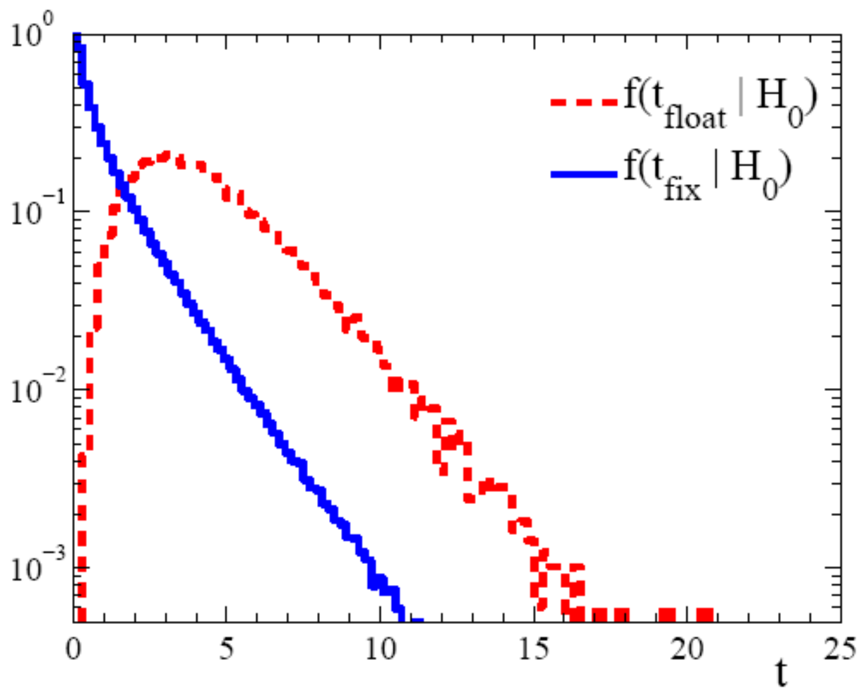
(Note m does not appear in the $\mu = 0$ model.)

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

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.

Approximate correction for LEE

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

Gross and Vitells show the p-values are approximately related by

$$p_{\mathrm{global}} \approx p_{\mathrm{local}} + \langle N(c) \rangle$$

where $\langle N(c) \rangle$ is the mean number "upcrossings" of $t_{\mathrm{fix}} = -2 \ln \lambda$ in the fit range based on a threshold

$$c = t_{\mathrm{fix,obs}} = Z_{\mathrm{local}}^2$$

and where $Z_{\mathrm{local}} = \Phi^{-1}(1 - p_{\mathrm{local}})$ is the local significance.

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).

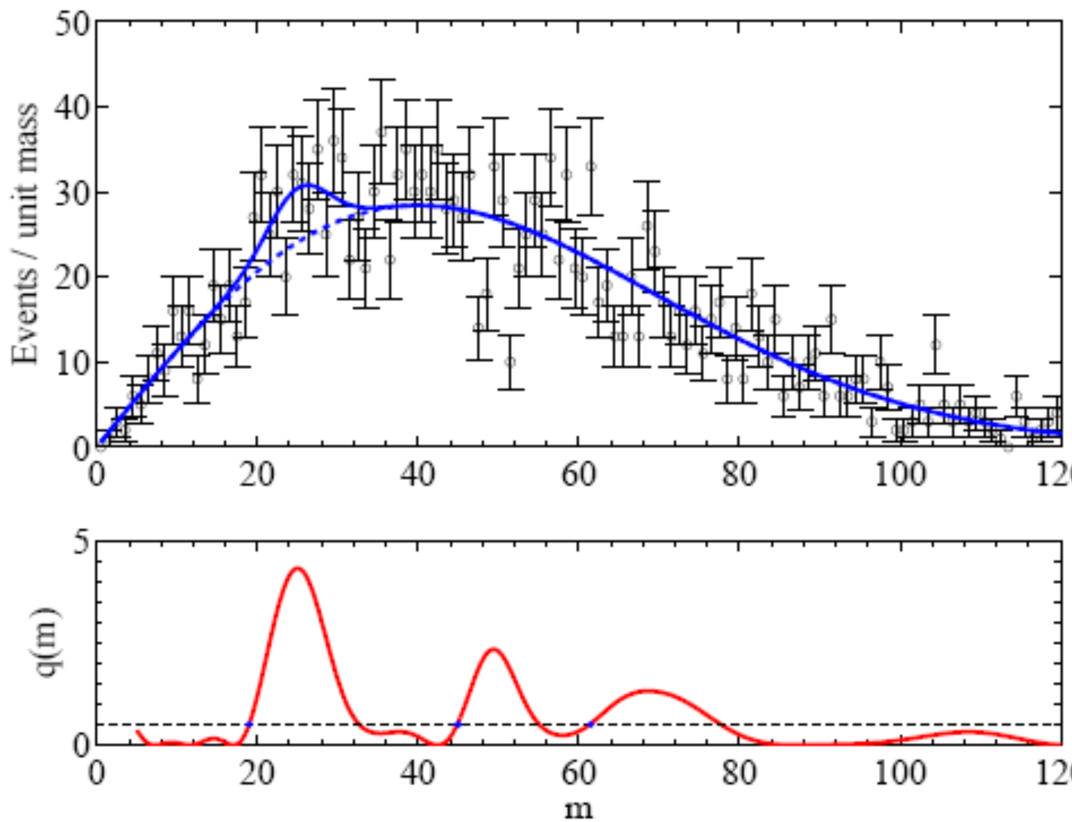
Upcrossings of $-2\ln L$

The Gross-Vitells formula for the trials factor requires $\langle N(c)\rangle$, the mean number "upcrossings" of $t_{\rm fix} = -2\ln\lambda$ in the fit range based on a threshold $c = t_{\rm fix} = Z_{\rm fix}^2$.

$\langle N(c)\rangle$ can be estimated from MC (or the real data) using a much lower threshold c_0:

$$\langle N(c)\rangle \approx \langle N(c_0)\rangle e^{-(c-c_0)/2}$$

In this way $\langle N(c)\rangle$ can be estimated without need of large MC samples, even if the the threshold c is quite high.

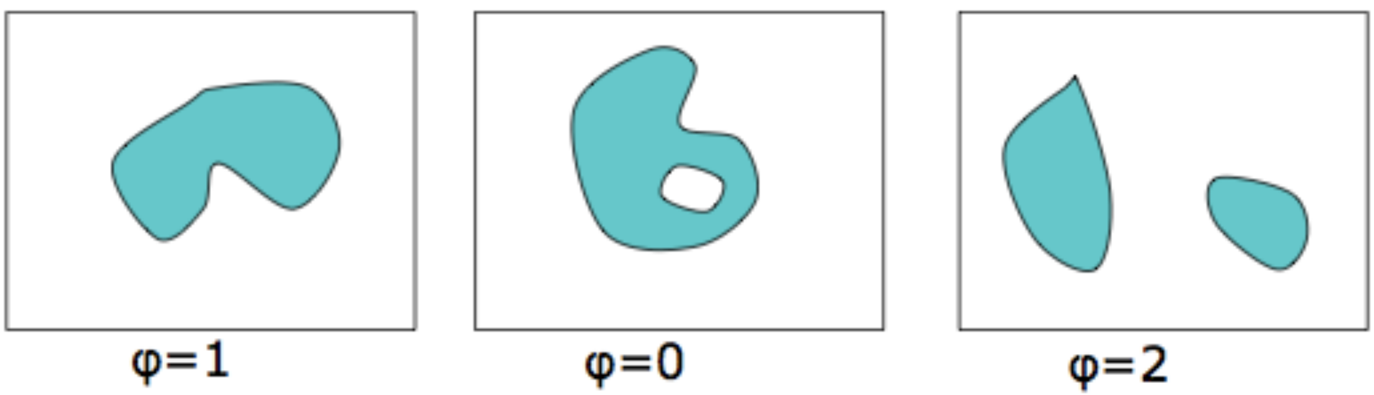


Multidimensional look-elsewhere effect

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, ...

Summary on Look-Elsewhere Effect

Remember the Look-Elsewhere Effect is when we test a single model (e.g., SM) with multiple observations, i..e, in mulitple places.

Note there is no look-elsewhere effect when considering exclusion limits. There we test specific signal models (typically once) and say whether each is excluded.

With exclusion there is, however, the also problematic issue of testing many signal models (or parameter values) and thus excluding some for which one has little or no sensitivity.

Approximate correction for LEE should be sufficient, and one should also report the uncorrected significance.

"There's no sense in being precise when you don't even know what you're talking about." — John von Neumann

Why 5 sigma?

Common practice in HEP has been to claim a discovery if the p-value of the no-signal hypothesis is below 2.9×10^{-7}, corresponding to a significance $Z = \Phi^{-1}(1 - p) = 5$ (a 5σ effect).

There a number of reasons why one may want to require such a high threshold for discovery:

The "cost" of announcing a false discovery is high.

Unsure about systematics.

Unsure about look-elsewhere effect.

The implied signal may be a priori highly improbable (e.g., violation of Lorentz invariance).

Why 5 sigma (cont.)?

But the primary role of the *p*-value is to quantify the probability that the background-only model gives a statistical fluctuation as big as the one seen or bigger.

It is not intended as a means to protect against hidden systematics or the high standard required for a claim of an important discovery.

In the processes of establishing a discovery there comes a point where it is clear that the observation is not simply a fluctuation, but an "effect", and the focus shifts to whether this is new physics or a systematic.

Providing LEE is dealt with, that threshold is probably closer to 3σ than 5σ.

Summary

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 for 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.

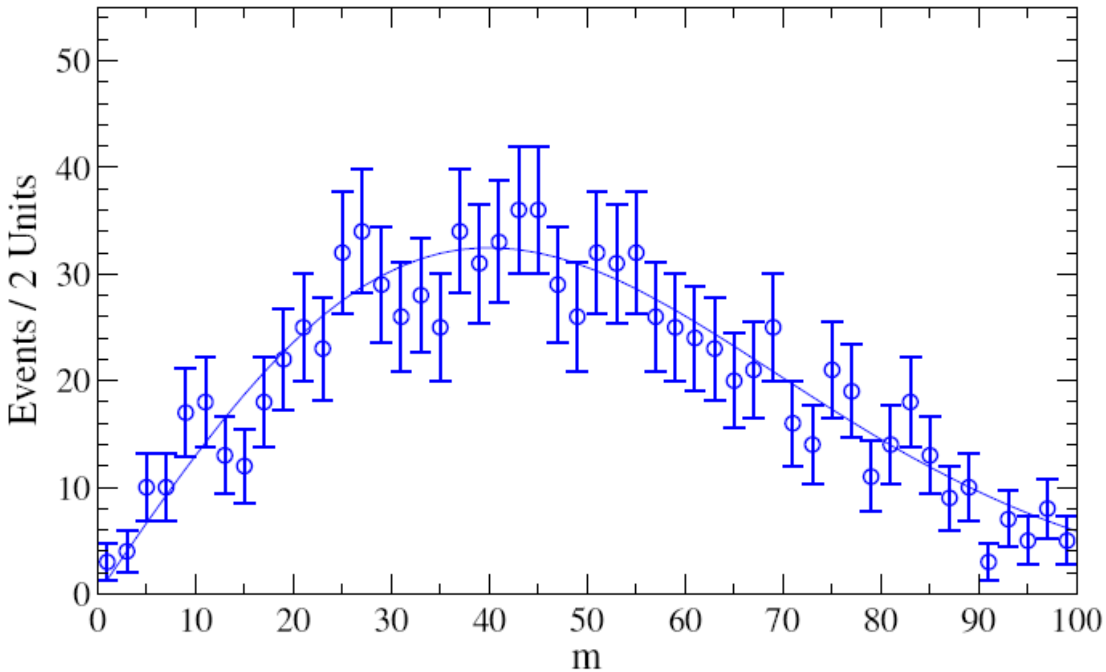
Implementation in RooStats (KC).

Thanks to Louis Fayard, Nancy Andari, Francesco Polci, Marumi Kado for their observations related to allowing a negative estimator for μ.

Extra slides

Example: Shape analysis

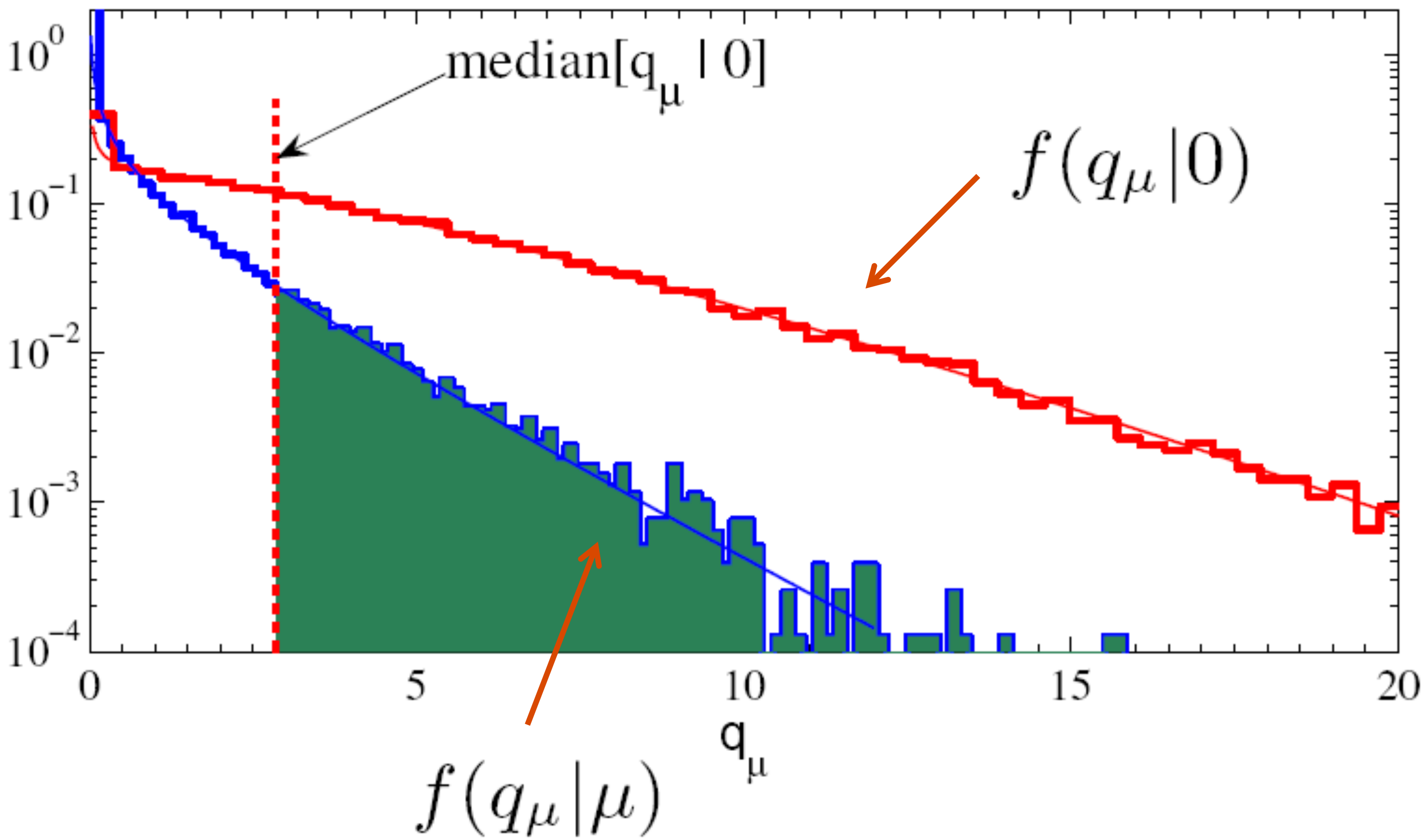
Look for a Gaussian bump sitting on top of:



$$L(\mu, \theta) = \prod_{i=1}^{N} \frac{(\mu s_i + \theta f_{\mathrm{b},i})^{n_i}}{n_i!} e^{-(\mu s_i + \theta f_{\mathrm{b},i})}$$

Monte Carlo test of asymptotic formulae

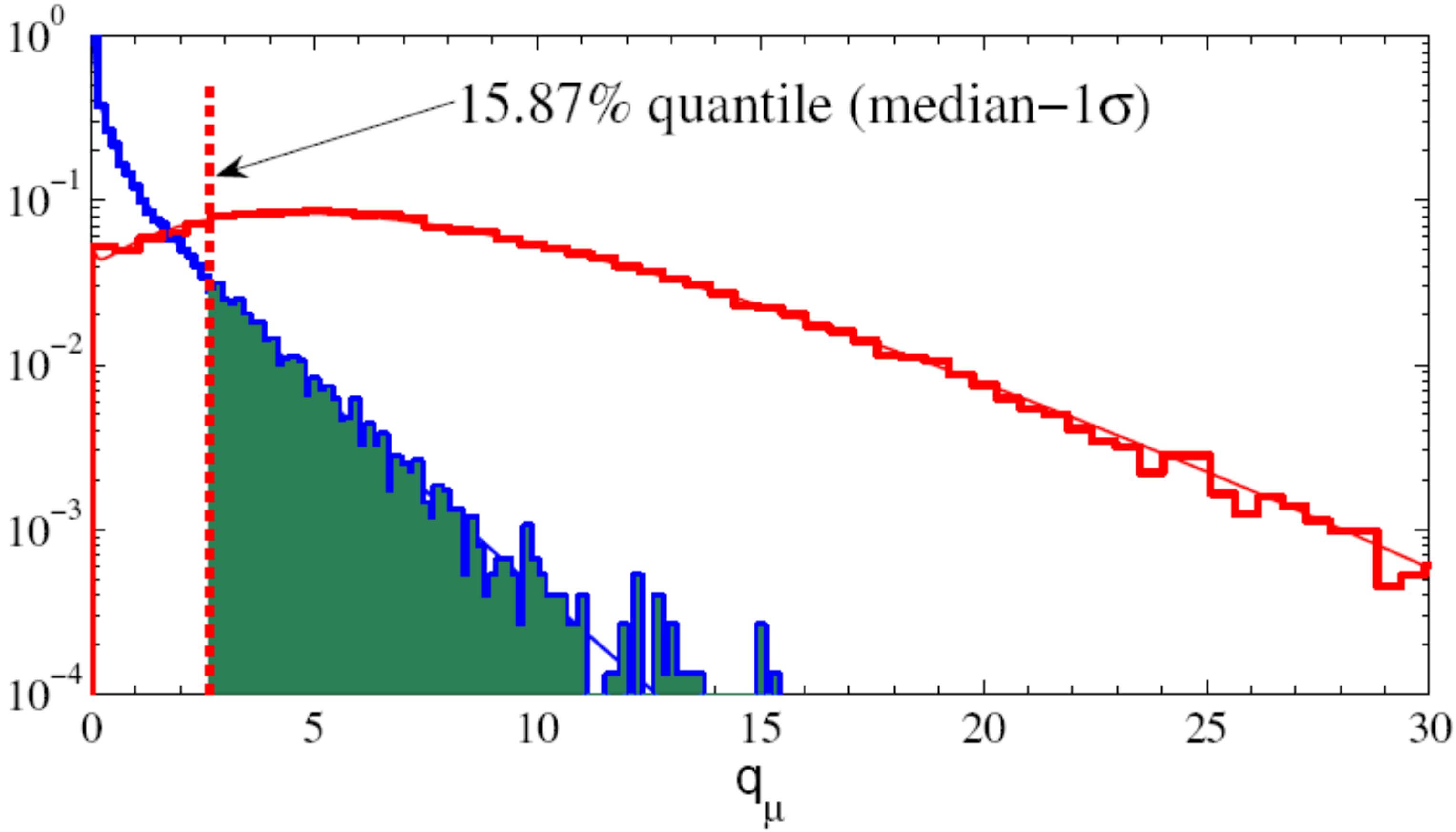
Distributions of q_μ here for μ that gave $p_\mu = 0.05$.



Using $f(q_\mu|0)$ to get error bands

We are not only interested in the median$[q_\mu|0]$; we want to know how much statistical variation to expect from a real data set.

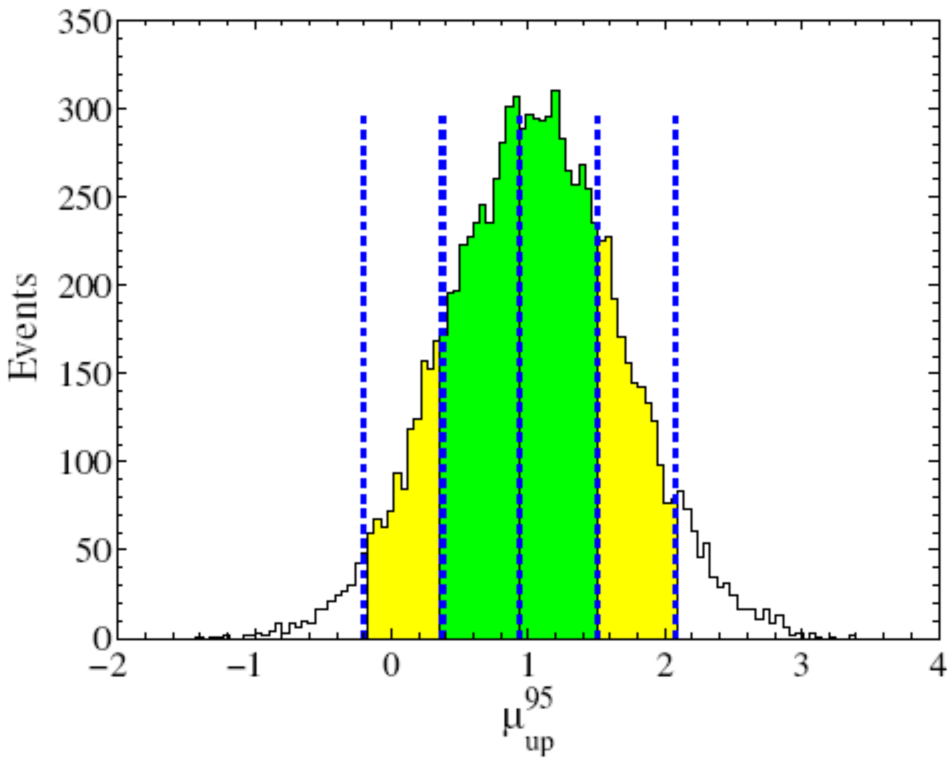
But we have full $f(q_\mu|0)$; we can get any desired quantiles.



Distribution of upper limit on μ

$\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands from MC;

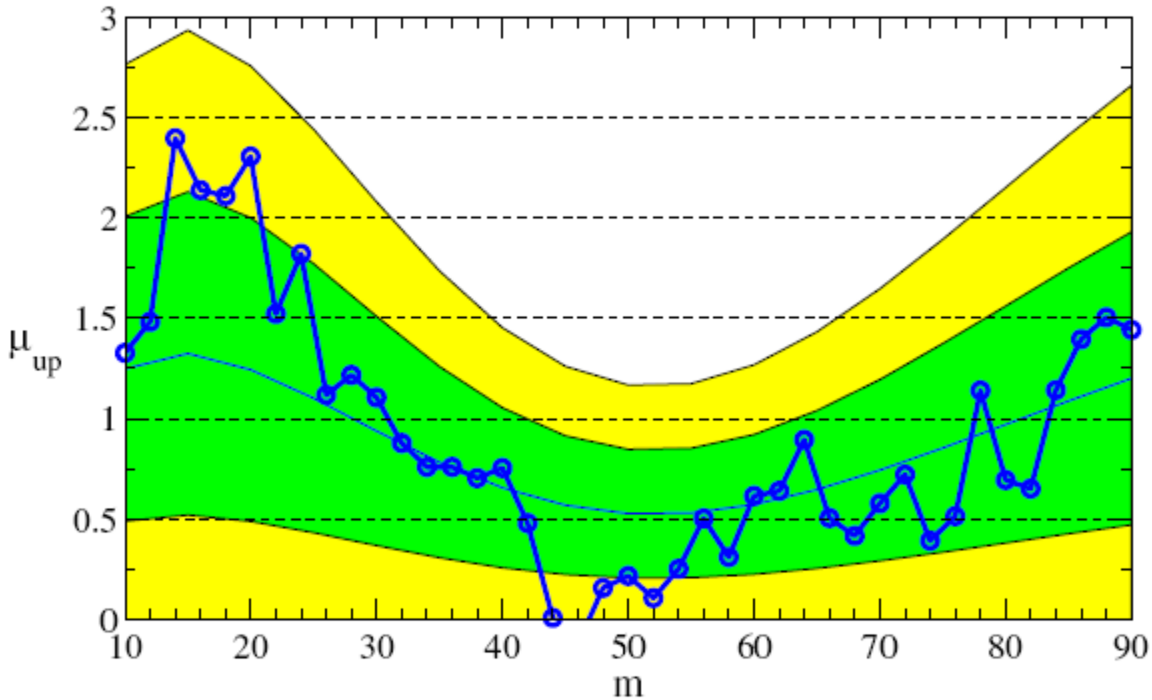
Vertical lines from asymptotic formulae



Limit on μ versus peak position (mass)

$\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands from asymptotic formulae;

Points are from a single arbitrary data set.



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

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

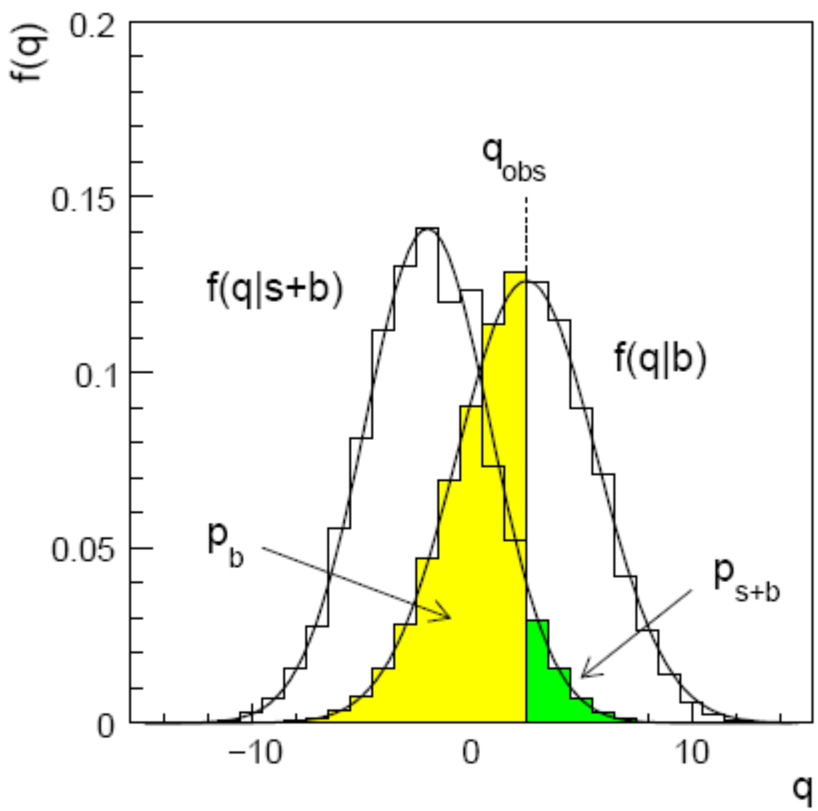
i.e. q is Gaussian distributed with mean and variance of

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

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

Example with L_{s+b}/L_b

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



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