Two developments in discovery tests:
use of weighted Monte Carlo events and an improved measure of experimental sensitivity

Progress on Statistical Issues in Searches
SLAC, 4-6 June, 2012

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

Outline

Two issues of practical importance in recent LHC analyses:

1) In many searches for new signal processes, estimates of rates of some background components often based on Monte Carlo with weighted events. Some care (and assumptions) are required to assess the effect of the finite MC sample on the result of the test.

2) A measure of discovery sensitivity is often used to plan a future analysis, e.g., $s/\sqrt{b}$, gives approximate expected discovery significance (test of $s = 0$) when counting $n \sim$ Poisson($s+b$). A measure of discovery significance is proposed that takes into account uncertainty in the background rate.

Using MC events in a statistical test

Prototype analysis – count n events where signal may be present:

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

s = expected events from nominal signal model (regard as known)
b = expected background (nuisance parameter)
μ = strength parameter (parameter of interest)

Ideal – constrain background b with a data control measurement m, scale factor τ (assume known) relates control and search regions:

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

Reality – not always possible to construct data control sample, sometimes take prediction for b from MC.

From a statistical perspective, can still regard number of MC events found as $m \sim \text{Poisson}(\tau b)$ (really should use binomial, but here Poisson good approx.) Scale factor is $\tau = L_{\text{MC}}/L_{\text{data}}$.

MC events with weights

But, some MC events come with an associated weight, either from generator directly or because of reweighting for efficiency, pile-up.

Outcome of experiment is: $n, m, w_1, \ldots, w_m$

How to use this info to construct statistical test of μ?

"Usual" (?) method is to construct an estimator for b:

$$\hat{b} = \frac{1}{\tau} \sum_{i=1}^{m} w_i \qquad\qquad \hat{\sigma}_{\hat{b}}^2 = \frac{1}{\tau^2} \sum_{i=1}^{m} w_i^2$$

and include this with a least-squares constraint, e.g., the χ^2 gets an additional term like

$$\frac{(b - \hat{b})^2}{\hat{\sigma}_{\hat{b}}^2}$$

Case where m is small (or zero)

Using least-squares like this assumes $\hat{b}$ ~ Gaussian, which is OK for sufficiently large m because of the Central Limit Theorem.

But $\hat{b}$ may not be Gaussian distributed if e.g.

m is very small (or zero),

the distribution of weights has a long tail.

Hypothetical example:

$m = 2, w_1 = 0.1307, w_2 = 0.0001605,$

$\hat{b} = 0.0007 \pm 0.0030$

$n = 1$ (!)

Correct procedure is to treat m ~ Poisson (or binomial). And if the events have weights, these constitute part of the measurement, and so we need to make an assumption about their distribution.

Constructing a statistical test of μ

As an example, suppose we want to test the background-only hypothesis (μ=0) using the profile likelihood ratio statistic (see e.g. CCGV, EPJC 71 (2011) 1554, arXiv:1007.1727),

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

where

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

From the observed value of q_0, the p-value of the hypothesis is:

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

So we need to know the distribution of the data (n, m, w_1,..., w_m), i.e., the likelihood, in two places:

1) to define the likelihood ratio for the test statistic

2) for $f(q_0|0)$ to get the p-value

Normal distribution of weights

Suppose $w \sim$ Gauss (ω, σ_w). The full likelihood function is

$$L(\mu, b, \omega, \sigma_w) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \frac{(\tau b/\omega)^m}{m!} e^{-\tau b/\omega} \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_w} e^{(w_i - \omega)^2 / 2\sigma_w^2}$$

The log-likelihood can be written:

$$\begin{aligned} \ln L(\mu, b, \omega, \sigma_w) &= n \ln(\mu s + b) - (\mu s + b) + m \ln(\tau b/\omega) - \tau b/\omega \\ &- m \ln \sigma_w - \frac{m\omega^2}{2\sigma_w^2} + \frac{\omega}{\sigma_w^2} \sum_{i=1}^{m} w_i - \frac{1}{2\sigma_w^2} \sum_{i=1}^{m} w_i^2 + C \end{aligned}$$

Only depends on weights through: $S_1 = \sum_{i=1}^{m} w_i \, , \quad S_2 = \sum_{i=1}^{m} w_i^2 \, .$

Log-normal distribution for weights

Depending on the nature/origin of the weights, we may know:

$w(x) \geq 0$,

distribution of w could have a long tail.

So $w \sim$ log-normal could be a more realistic model.

I.e, let $l = \ln w$, then $l \sim$ Gaussian(λ, σ_l), and the log-likelihood is

$$\begin{aligned} \ln L(\mu, b, \lambda, \sigma_l) &= n \ln(\mu s + b) - (\mu s + b) + m \ln(\tau b/\omega) - \tau b/\omega \\ &- m \ln \sigma_l - \frac{m\lambda^2}{2\sigma_l^2} + \frac{\lambda}{\sigma_l^2} \sum_{i=1}^{m} l_i - \frac{1}{2\sigma_l^2} \sum_{i=1}^{m} l_i^2 \,. \end{aligned}$$

where $\lambda = E[l]$ and $\omega = E[w] = \exp(\lambda + \sigma_l^2/2)$.

Need to record n, m, $\Sigma_i \ln w_i$ and $\Sigma_i \ln^2 w_i$.

Normal distribution for $\hat{b}$

For $m > 0$ we can define the estimator for b

$$\hat{b} = \frac{1}{\tau} \sum_{i=1}^{m} w_i \qquad \hat{\sigma}_{\hat{b}}^2 = \frac{1}{\tau^2} \sum_{i=1}^{m} w_i^2$$

If we assume $\hat{b}$ ~ Gaussian, then the log-likelihood is

$$\ln L(\mu, b) = n \ln(\mu s + b) - (\mu s + b) - \frac{1}{2} \frac{(b - \hat{b})^2}{\hat{\sigma}_{\hat{b}}^2}$$

Important simplification: L only depends on parameter of interest μ and single nuisance parameter b.

Ordinarily would only use this Ansatz when Prob(m=0) negligible.

Toy weights for test of procedure

Suppose we wanted to generate events according to

$$f(x) = \frac{e^{-x/\xi}}{\xi(1 - e^{-a/\xi})} , \quad 0 \leq x \leq a .$$

Suppose we couldn't do this, and only could generate *x* following

$$g(x) = \frac{1}{a} , \quad 0 \leq x \leq a$$

and for each event we also obtain a weight

$$w(x) = \frac{f(x)}{g(x)} = \frac{a}{\xi} \frac{e^{-x/\xi}}{1 - e^{-a/\xi}}$$

In this case the weights follow:

$$p(w) = \frac{\xi}{aw}$$

$$w_{\rm min} \; \leq \; w \; \leq w_{\rm max}$$

Two sample MC data sets

Suppose $n = 17$, $\tau = 1$, and

case 1:

$a = 5$, $\xi = 25$

$m = 6$

Distribution of w narrow

weight w	$\ln w$
0.9684	-0.0320
0.9217	-0.0816
1.0238	0.0235
1.0063	0.0063
0.9709	-0.0295
1.0813	0.0782

case 2:

$a = 5$, $\xi = 1$

$m = 6$

Distribution of w broad

weight w	$\ln w$
0.1934	-1.6429
0.0561	-2.8809
0.7750	-0.2548
0.5039	-0.6853
0.2059	-1.580
3.0404	1.1120

Testing $\mu = 0$ using q_0 with $n = 17$

case 1:
$a = 5, \xi = 25$
$m = 6$
Distribution of
w is narrow

Likelihood used to define q_0	Distribution of w for $f(q_0\|0)$	Significance Z to reject $\mu = 0$
$w \sim$ normal	normal	2.287
$w \sim$ normal	$1/w$	2.268
$w \sim$ log-normal	log-normal	2.301
$w \sim$ log-normal	$1/w$	2.267
$\hat{b} \sim$ normal	normal	2.289
$\hat{b} \sim$ normal	$1/w$	2.224

If distribution of weights is narrow, then all methods result in a similar picture: discovery significance $Z \sim 2.3$.

Testing $\mu = 0$ using q_0 with $n = 17$ (cont.)

case 2:
$a = 5$, $\xi = 1$
$m = 6$
Distribution of w is broad

Likelihood used to define q_0	Distribution of w for $f(q_0\|0)$	Significance Z to reject $\mu = 0$
$w \sim$ normal	normal	2.163
$w \sim$ normal	$1/w$	1.308
$w \sim$ log-normal	log-normal	0.863
$w \sim$ log-normal	$1/w$	0.983
$\hat{b} \sim$ normal	normal	1.788
$\hat{b} \sim$ normal	$1/w$	1.387

If there is a broad distribution of weights, then:

1) If true $w \sim 1/w$, then assuming $w \sim$ normal gives too tight of constraint on b and thus overestimates the discovery significance.

2) If test statistic is sensitive to tail of w distribution (i.e., based on log-normal likelihood), then discovery significance reduced.

Best option above would be to assume $w \sim$ log-normal, both for definition of q_0 and $f(q_0|0)$, hence $Z = 0.863$.

Summary on weighted MC

Treating MC data as "real" data, i.e., $n \sim$ Poisson, incorporates the statistical error due to limited size of sample.

Then no problem if zero MC events observed, no issue of how to deal with 0 ± 0 for background estimate.

If the MC events have weights, then some assumption must be made about this distribution.

If large sample, Gaussian should be OK,

if sample small consider log-normal.

See draft note for more info and also treatment of weights $= \pm 1$ (e.g., MC@NLO).

www.pp.rhul.ac.uk/~cowan/stat/notes/weights.pdf

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($x > x_{obs}$ | $s = 0$),:

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

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),

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

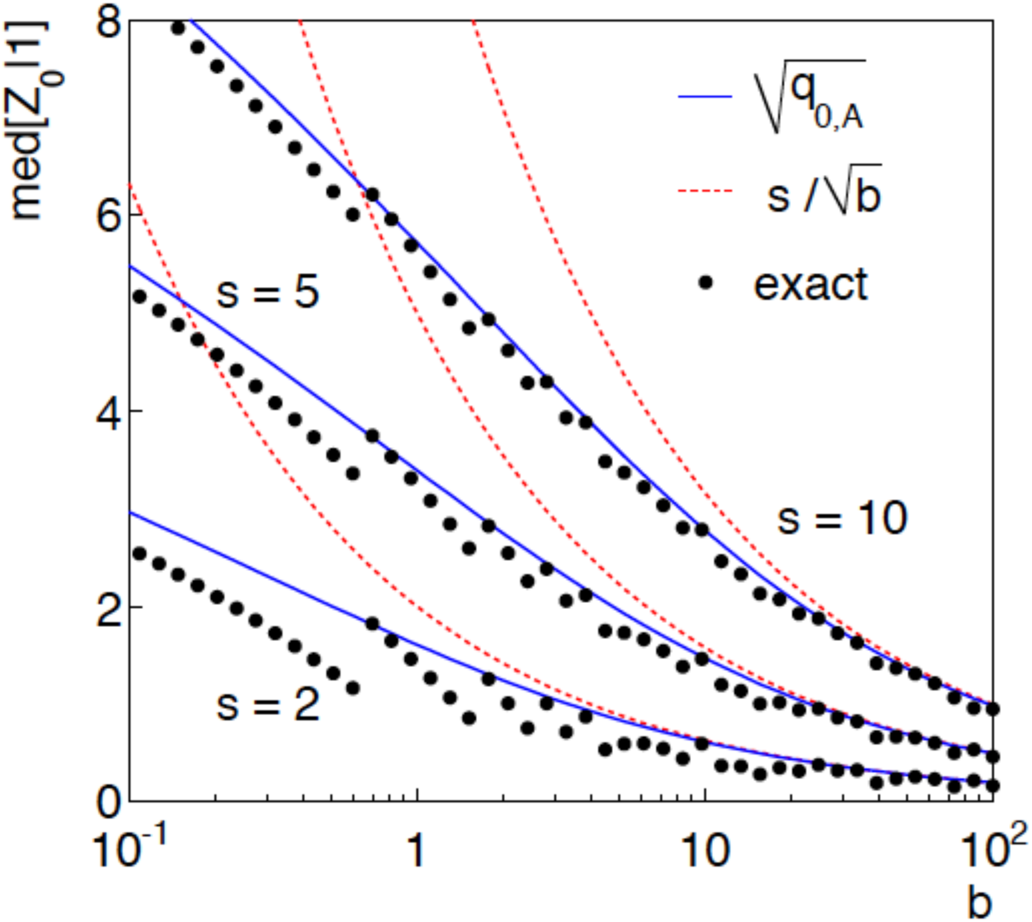
To find median[$Z|s$], let $n \rightarrow s + b$ (i.e., the Asimov data set):

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

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):

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

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:

$$n \rightarrow s + b$$

$$m \rightarrow \tau b$$

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

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

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

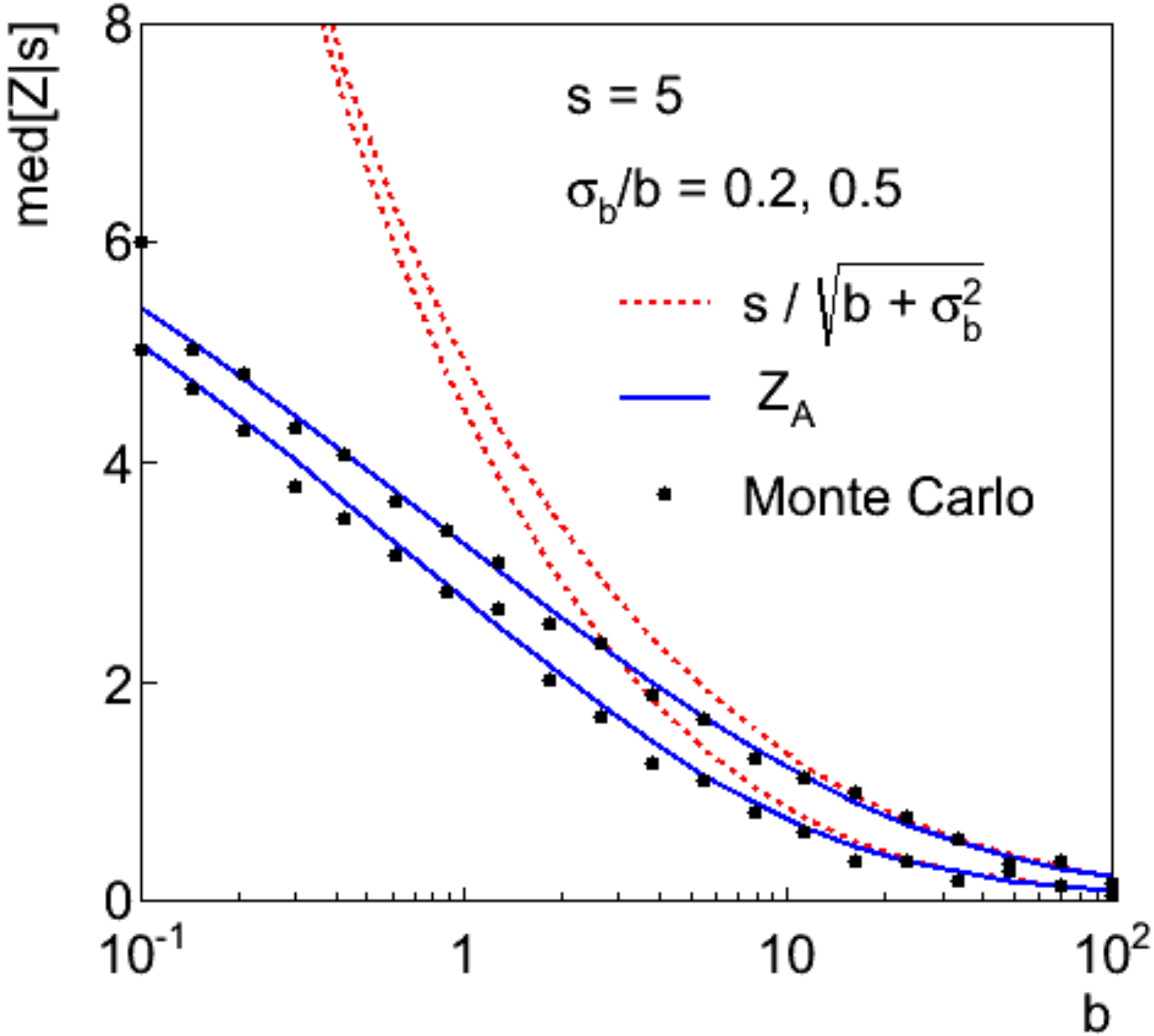
Limiting cases

Expanding the Asimov formula in powers of s/b and $\sigma_b{}^2/b$ (= $1/\tau$) 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: $s = 5$



Summary on discovery sensitivity

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

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

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 (check this).

Extra slides

Ingredients for profile likelihood ratio

To construct the profile likelihood ratio we need the estimators:

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

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

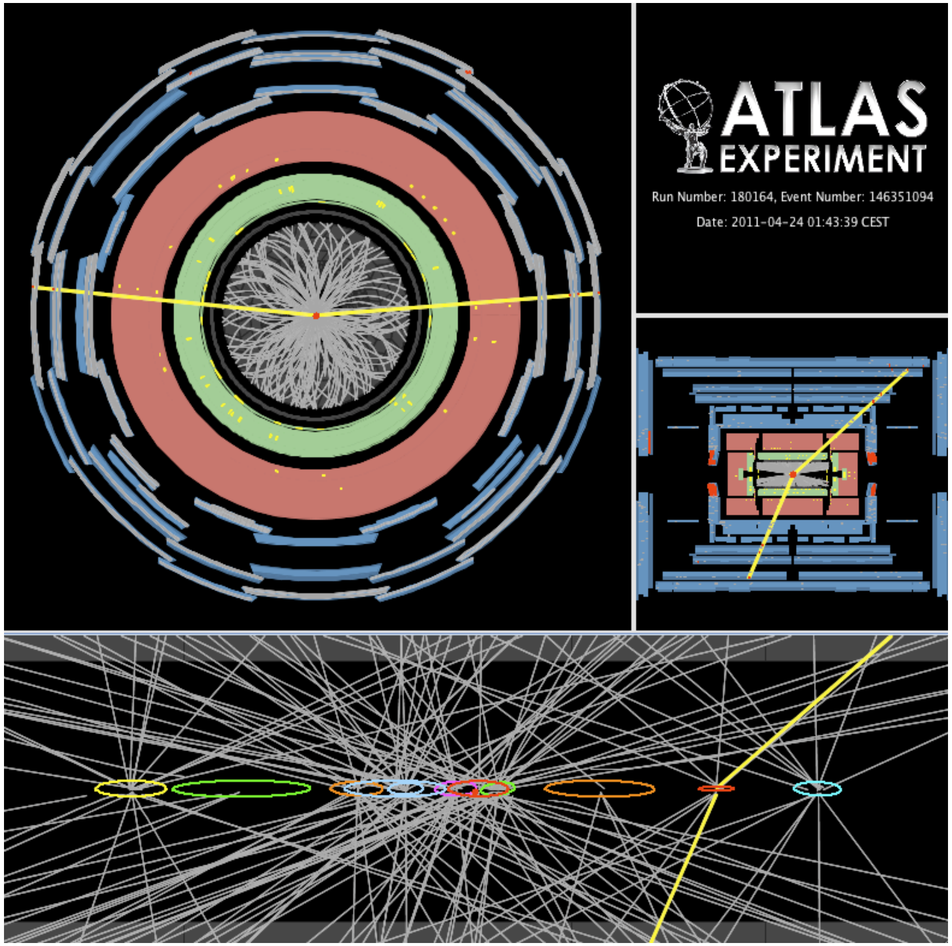
$$\hat{\hat{b}}(s) = \frac{n + m - (1+\tau)s + \sqrt{(n + m - (1+\tau)s)^2 + 4(1+\tau)sm}}{2(1+\tau)} \,.$$

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

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

MC weights due to pile-up

Each pp bunch crossing at the LHC results in a number of pp collisions (11 in the one below); this is "pile-up".



The effect of pile-up is included in MC sim-ulations using a best guess for the rate, which depends e.g. on beam intensity.

If this guess turns out to be incorrect, the MC events are reweighted to correct the distribution of the number of collisions per bunch crossing.

Using sensitivity to optimize a cut

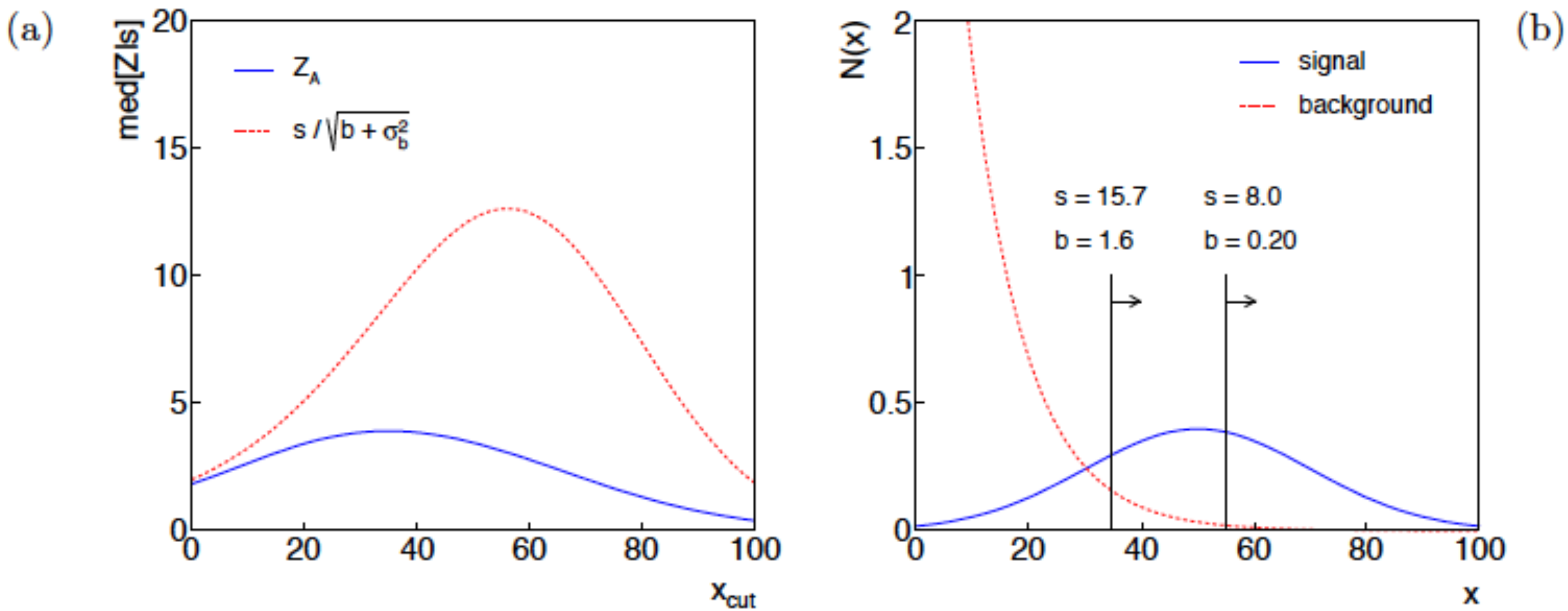


Figure 1: (a) The expected significance as a function of the cut value x_{cut}; (b) the distributions of signal and background with the optimal cut value indicated.

Distributions of q_0

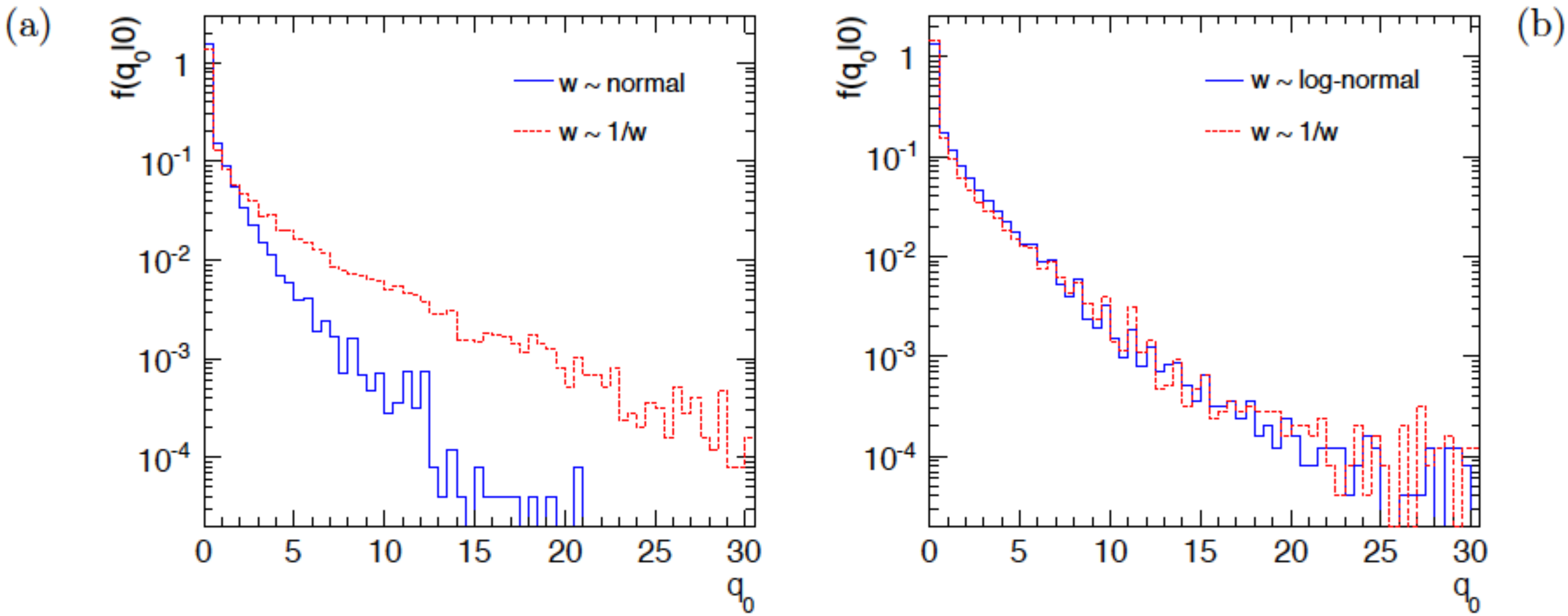


Figure 2: Distributions of the statistic q_0 based the profile likelihood using (a) a normal model for the weights and (b) on a log-normal model. In each plot the curves are shown representing two assumptions for the distribution of weights: the same as used to define q_0 (normal or log-normal) and the $1/w$ distribution.