Statistical tests with weighted Monte Carlo events

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

ATLAS/CMS Statistics Meeting
EVO, 26 September, 2011

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

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:

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

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

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

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.

Suppose e.g.:

$m = 2$, $w_1 = 0.1$, $w_2 = 0.0001$,

$\hat{b}$ = small

$n = 1$ (!)

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

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

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

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

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

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

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:

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

$$w_{\rm min} \le w \le w_{\rm max}$$

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 $Z = 0.863$.

Distributions of q_0

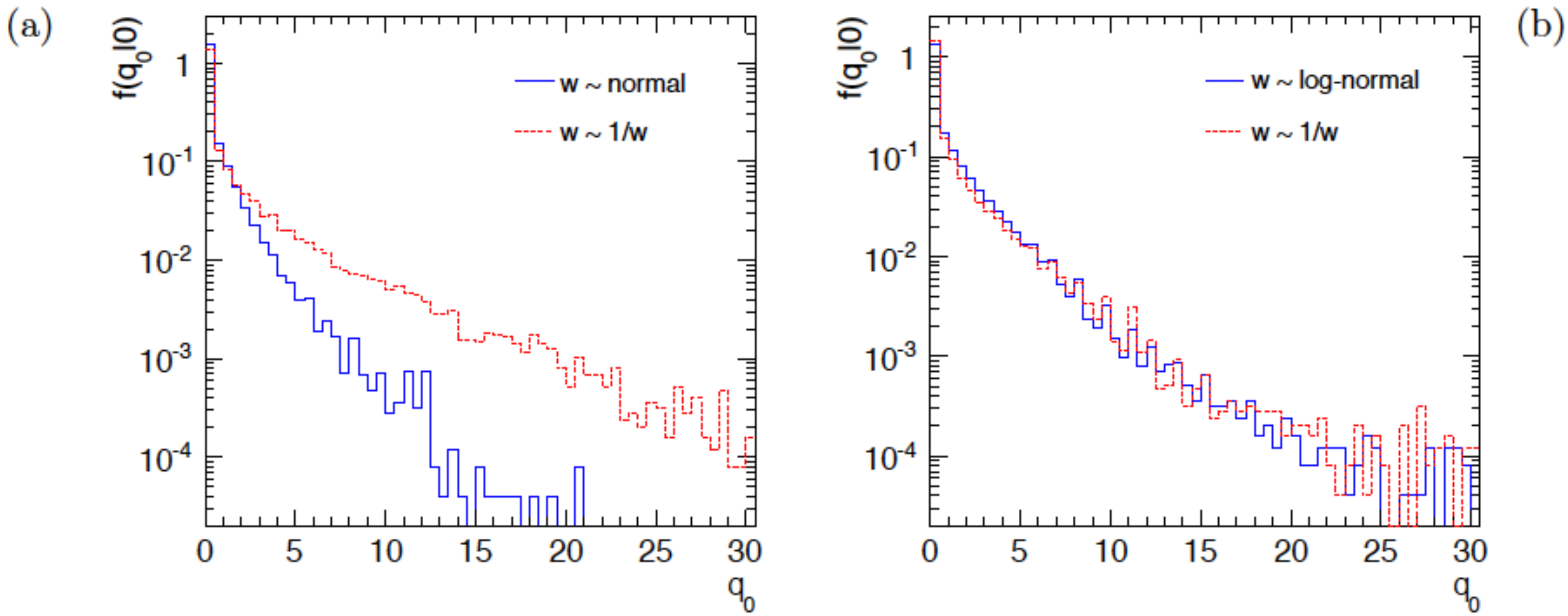

Figure 2: Distributions of the statistic q_0 based the profile likelihood using (a) a normal model for the weights and (b) on a log-normal model. In each plot the curves are shown representing two assumptions for the distribution of weights: the same as used to define q_0 (normal or log-normal) and the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

Summary

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 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`