Asimov estimate for discovery significance in counting experiment

ATLAS Statistics Forum
CERN/Phone, 15 February 2011

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

Outline

Today two brief comments:

1) At 5 October 2010 Statistics Forum meeting we discussed presentation of error bars on plots. I have written this up as a note and attached to today's agenda (DRAFT).

2) The "Asimov Paper" (Cowan, Cranmer, Gross, Vitells, arXiv: 1007.1727) is now EPJC 71 (2011) 1-19. It contains a formula for the expected discovery significance for Poisson data – I will make a brief comment on this.

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

Consider the case where we observe n events, model as following Poisson distribution with mean $s + b$.

Here 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 $s + b$, $n \rightarrow x \sim$ 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

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

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 $\frac{\partial \ln L}{\partial s} = 0$

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

Approximate Poisson significance (continued)

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

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

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

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

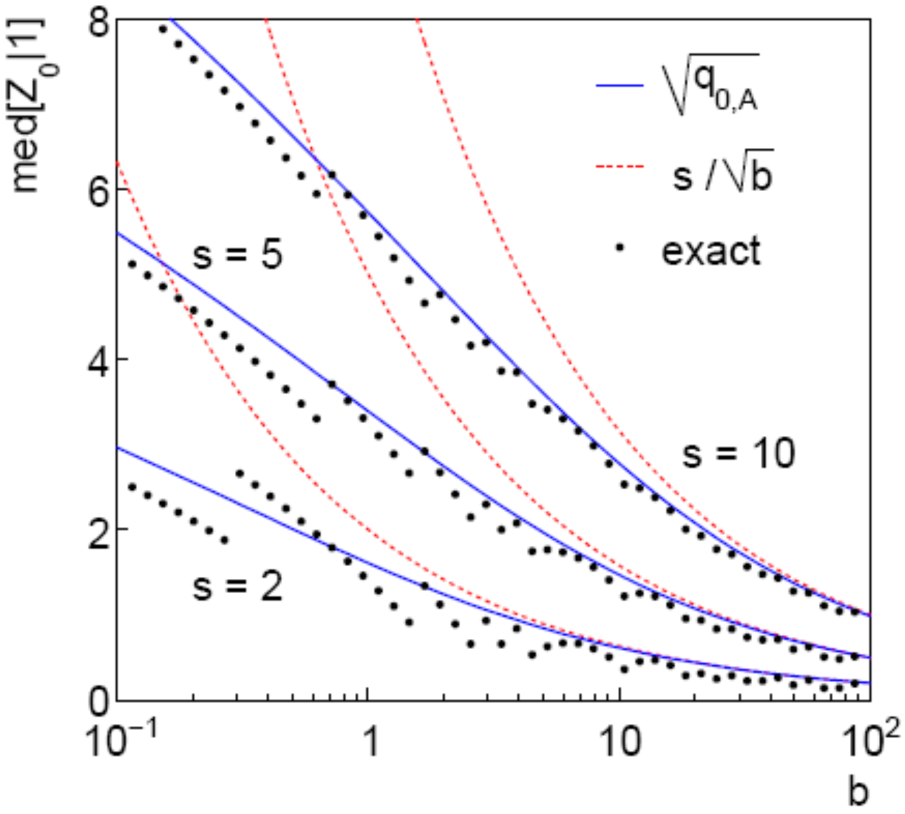
To find median[$Z_0|s+b$], let $n \to s + b$ (i.e., the Asimov data set):

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

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

$n \sim$ 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 $s \ll b$.